

Global Observation and Knowledge Infrastructures: A Study of Collective Inquiry across Heterogeneous Relational Systems

Wanhong HUANG
huangwanhong@serendip.ngo

Abstract

International observation systems reveal a consequential possibility: scientific information can be generated, validated, and circulated across institutional and political boundaries without requiring a single authority over all observations or a uniform distribution of capacities. Yet the existence of data-sharing protocols does not by itself establish shared understanding, mutual trust, legitimate reuse, or equitable participation in the production of knowledge. This discussion paper investigates the technical, epistemic, and institutional conditions under which *global observation and knowledge infrastructures* might be constructed and sustained. Drawing on meteorological and oceanographic cooperation, it distinguishes observational coverage, data availability, interoperability, interpretation, and collective authority. It proposes a preliminary federated research architecture connecting situated observation nodes, provenance and quality records, semantic and model translation, distributed computation, access governance, correction, and contestation. It treats mathematical, physical, social, and normative languages as historically formed manifestations and practices within generative-relational systems, distinguishing their conditional formal invariance from claims to universally identical representation. It further investigates why actors situated in different generative and institutional histories may interpret shared evidence differently, and proposes a bounded research vocabulary for mutually workable norms, continued disagreement, and rule revision without assuming epistemic consensus. The paper conceptualizes trust as a historically constituted relation supported in part by verifiable commitments, reciprocal inquiry, accountable institutions, and effective remedy rather than as a property supplied by cryptography or open data alone. It further specifies prospective services for common knowledge creation and circulation, generalized translation of heterogeneous generative-relational contexts, and traceable transformations of distinct forms of value. A staged validation design differentiates synthetic policy gates, cross-node scientific comparison, productive-capacity trials, and site-dependent ecological and institutional review. A cross-layer assessment profile separates technical interchange, scientific warrant, claimed competent authority, and effective contestation, without deriving legal or ethical legitimacy from technical success. It also examines epistemic and infrastructural asymmetries, the sensitivity of ecological and community data, the incompleteness of model-based assimilation, and the risks of concentrating interpretive authority in highly capable AI systems. Global weather and ocean observation provide instructive but bounded cases, not evidence that all knowledge should be made public or placed under centralized control. The proposed programme remains exploratory and separates conceptual architecture from implemented infrastructure, independent empirical assessment, and normative justification.

Keywords: global observation; knowledge infrastructure; epistemic interoperability; federated systems; data provenance; data assimilation; international cooperation; epistemic bias; common knowledge creation; generalized translation; heterogeneous value circulation; generalized language; reality registers; normative coordination; trust; governance; ocean observation.

Discussion Paper Note

This independent research discussion paper investigates conditions for constructing global observation and knowledge infrastructures through federated scientific capacities, common knowledge creation and circulation, and generalized mediation among heterogeneous relational systems. The manuscript contains fourteen main sections and seven appendices; its architectural profiles, translation models, mathematical statements, and proposed evaluation protocols have undergone five bounded internal revision stages. They remain research constructions rather than implemented or independently validated international services. In particular, technical compatibility, epistemic warrant, claimed competent authority, and opportunities for contestation are separate assessments. Neither authenticated records, participant agreement, formal notation, nor synthetic tests alone establish legal validity, effective remedy, environmental benefit, or field performance. The paper reports no original worldwide observation campaign, deployed global federation, binding legal determination, marine ecological impact assessment, or assessed multinational implementation. Existing scientific programmes are discussed as bounded precedents and are not presented as endorsing this proposal.

Notices

Publication status. Independent preliminary discussion-paper draft v2.9. All five planned internal revision stages are complete, including a selected source-to-claim review, formal and architectural checks, synthetic test cases, and an editorial preflight. Completion of that internal process does not amount to peer review, full-text verification of all cited research, independent mathematical proof checking, domain-engineering certification, legal advice, ecological assessment, or participation and consent by potentially affected communities. These matters remain subject to appropriately scoped independent investigation.

Licence. Except where otherwise indicated, copyright © 2026 Wanhong Huang. Original material is licensed under Creative Commons Attribution–NonCommercial 4.0 International (CC BY-NC 4.0): <https://creativecommons.org/licenses/by-nc/4.0/>. The rights of third-party materials remain with their holders.

AI-assisted preparation. OpenAI’s ChatGPT assisted with dialogue, initial drafting, reference discovery, and LaTeX preparation. The human author directs the research and assumes responsibility for factual verification, interpretation, and dissemination.

Scope. Descriptions of public scientific and institutional programmes are distinguished from hypothetical architecture and simulated exercises. Proposed decision gates and legal-source profiles are not authorization instruments, and no causal policy impacts or worldwide deployments are claimed.

1 Introduction

The construction of a global scientific infrastructure is frequently described as a problem of connecting data. Sensors acquire measurements, organizations publish records, repositories distribute files, models assimilate observations, and users transform model outputs into practical decisions. This description captures genuine engineering achievements, but it leaves open a more demanding question: under what technical, epistemic, and institutional conditions can geographically and politically distributed participants jointly produce *credible knowledge* without surrendering their capacity to observe, interpret, contest, and govern what is known? The present inquiry investigates this question through the prospective formation of global observation and knowledge infrastructures. Its principal object is the architecture

of cooperation among heterogeneous observational systems, interpretive practices, computational facilities, institutions, and affected communities. It examines the enabling conditions of reliable information exchange while resisting the assumption that more extensive connectivity by itself produces a common understanding of the world.

The question begins with a practical tension. Some phenomena can be investigated only through observations collected across extensive and differently governed regions. Atmospheric transport, large-scale ocean circulation, earthquakes, disease ecology, and many climate processes are not contained within the jurisdictions that finance their measurement. Oceanographic profiles gathered in one region can inform inferences about conditions elsewhere; data from one national weather service can improve the predictive basis of another. At the same time, the technical means of observation are distributed unevenly. States and research institutions differ in sensor coverage, calibration laboratories, staffing, access to ships and satellites, computing resources, and the ability to establish the categories through which observations will be interpreted. A common dataset may consequently carry a history of unequal capacities and contributions even when all its fields are technically accessible. A theory of global infrastructure has to account both for the benefits of shared inquiry and for the power associated with defining what counts as an observation, a legitimate interpretation, or a valid use.

Contemporary meteorological cooperation provides a substantial point of departure. The World Meteorological Organization’s Information System 2.0 (WIS2) entered its operational phase in January 2025. Its design makes use of open standards and Internet-based exchange, including a publish–subscribe model and distributed WIS2 nodes, rather than requiring every observation to be managed by a single central database [66, 68]. This arrangement is embedded in a longer history of meteorological coordination and a particular international institutional framework. The WMO Unified Data Policy, adopted in 2021, distinguishes *Core* data, which members shall exchange, from *Recommended* data, which they should exchange, and describes terms for international Earth-system information sharing [63]. The policy is instructive precisely because it does not establish that every observational asset or record is subject to identical access conditions. The case suggests a broader possibility: international scientific interdependence can motivate a common exchange system even when national responsibilities, ownership arrangements, and levels of willingness to share remain differentiated.

The ocean offers related but distinct lessons. The international Argo programme distributes observations from autonomous profiling floats through global data-management arrangements. Users can access real-time data and later, more extensively quality-controlled versions; corrections and revised records form part of the ongoing scientific process rather than evidence that an observation has acquired an immutable truth at the moment of collection [55]. The Global Ocean Observing System (GOOS), with supporting coordination such as OceanOPS, connects observing requirements, platform operations, common practices, metadata, and the flow of information among scientific and operational communities [58, 30]. The significance of these systems extends beyond the accumulation of measurements. They illustrate repeated cooperation in maintaining the *conditions of observation*: instrumentation, calibration, common metadata, quality assurance, data publication, and expert correction. At the same time, their differing mandates caution against treating meteorological exchange, float-data processing, and ocean-observing coordination as institutionally identical instances of a single global mechanism.

Other initiatives extend observation toward model integration and decision support. The European Digital Twin Ocean, supported by EDITO, seeks to bring observations, models, computational resources, and usable applications into a progressively integrated public knowledge environment. Official project material describes a continuing development programme and an ambition for a sustained, more fully operational infrastructure by 2030 [20]. It is therefore a relevant design trajectory, not proof that a complete digital representation of the ocean already exists. Likewise, the United Nations’ Early Warnings for All initiative explicitly distinguishes risk knowledge, observation and forecasting, warning communication, and preparedness and response. The resulting four-pillar account shows that globally shared observations have public value only through additional operational, communicative, institutional, and social arrangements [59]. An early warning can be physically accurate yet fail to reach the affected population, fail to be understood, or fail to prompt feasible protective action. In such circumstances the obstacle is not

64 simply an absent data point.

65 These experiences challenge one prevalent objection to global scientific infrastructure: that political
66 competition and persistent distrust render broad cooperation impossible. The objection captures genuine
67 difficulties, including strategic restrictions, uneven capabilities, conflicting interpretations, and contested
68 authority. Yet existing Earth-system arrangements demonstrate that repeated cross-border contributions
69 and common technical standards are possible without any general elimination of interstate competition.
70 This observation does not imply that the same degree of cooperation can simply be extended to every
71 scientific or commercial dataset. Meteorological and oceanographic networks have specific mandates,
72 long-developed professional communities, defined types of public benefit, and operational conventions.
73 Their relevance to other domains must be assessed through mechanisms rather than analogy alone. A
74 programme for global knowledge infrastructure should therefore inquire how particular interests in re-
75 ciprocal information, credible prediction, operational continuity, and shared public goods can support
76 cooperation, and under what conditions these interests fail to overcome asymmetry or conflict.

77 It is useful at the outset to distinguish five related but non-equivalent achievements: *observational*
78 *coverage*, *data availability*, *syntactic interoperability*, *epistemic interoperability*, and *legitimate coordi-*
79 *nation*. Coverage concerns what phenomena are sampled, where and under which observational lim-
80 itations. Availability concerns whether records can be discovered, obtained, and used by specified
81 participants under stated rules. Syntactic interoperability concerns formats, protocols, identifiers, and
82 machine-readable conventions. Epistemic interoperability concerns whether heterogeneous records and
83 models can be interpreted, compared, challenged, and translated without concealing their originating as-
84 sumptions or uncertainty. Legitimate coordination concerns who authorizes uses, whose interests count,
85 which disagreements can be raised, and what can be done after error or injury. Success on one dimen-
86 sion does not entail success on another. A shared file standard may carry incompatible measurements; a
87 technically sound model may embed disputed normative classifications; a high-quality public prediction
88 may depend upon proprietary computational facilities that other participating regions cannot meaning-
89 fully use.

90 The distinctions also expose a limitation in equating a global knowledge commons with the unre-
91 stricted release of all available records. Some Earth-system data are widely recognized as requiring
92 international exchange because rapid availability supports public forecasting and disaster preparedness.
93 Other information may identify individuals, reveal sensitive ecological sites, concern communities hold-
94 ing legitimate interests in interpretation and reuse, or be affected by public safety and security obligations.
95 The Findable, Accessible, Interoperable, and Reusable (FAIR) principles articulate important dimensions
96 of data stewardship and machine-actionable discovery [61]. The Collective Benefit, Authority to Control,
97 Responsibility, and Ethics (CARE) principles for Indigenous data governance emphasize complementary
98 questions about collective interests, control, accountability, and ethical use [6]. The intersection of these
99 approaches is not resolved by declaring a universal priority for access or restriction. Global cooperation
100 requires differentiated arrangements that make the purposes, permissions, boundaries, and consequences
101 of sharing visible and contestable. Otherwise, a purportedly public infrastructure can become an instru-
102 ment through which epistemic resources are extracted from poorly represented communities while the
103 capacity to interpret or benefit from them remains elsewhere.

104 The proposed research object is therefore broader than a global data catalogue. Observational records
105 participate in chains of calibration, correction, assimilation, modeling, interpretation, communication,
106 and institutional response. At each stage, additional assumptions become operative. A sensor returns a
107 signal under a particular configuration and set of environmental conditions. A preprocessing procedure
108 rejects some measurements, adjusts others, and attaches a quality category. An assimilation scheme com-
109 bines observations with a model and a representation of uncertainty; its output depends on the structure
110 and assumptions of the model as well as on the measurements. A downstream AI system may produce
111 a coherent linguistic or graphical interpretation, but coherence is not an independent validation of its
112 claims. For this reason, a knowledge infrastructure must preserve enough of the generative conditions
113 of its outputs for relevant users to identify, reproduce where feasible, and contest the transformations
114 involved. Reliable provenance is not a mere attribution field; it is part of the capacity to reconstruct why

115 a claim was formed.

116 The paper adopts a *relational* understanding of knowledge production.

117 In particular, the inquiry does not infer from the success of contemporary mathematical notation or
118 scientific practice that every conceivable observer must employ the same language. Mathematical struc-
119 tures and physical laws may nevertheless possess ontological reality independent of particular observers;
120 their ultimate status is an open question rather than something this paper denies. The present inquiry
121 concerns especially the relational generation of knowledge, proof practices, and scientific institutions,
122 and the manifold ways in which they become manifest. Their capacity for repeatable computation and
123 cross-institutional measurement is sustained in human settings by historically developed definitions, in-
124 struments, proofs, conventions, pedagogies, and cooperative practice. A formally specified identity such
125 as $1 + 1 = 2$ holds under the stated arithmetic interpretation; its warrant should be distinguished from the
126 physical or symbolic medium in which it is communicated and from an unsupported claim that all possi-
127 ble generative-relational systems must adopt the same representation. Another possible observer might
128 coordinate chemical or embodied signs around comparable regularities. This is a conceptual possibility,
129 not empirical evidence about extraterrestrial cognition. Section 2.4 develops the associated distinctions
130 among bracketed ontology, relational efficacy, and manifestation, and Section 6.2 examines their con-
131 sequences for translation. Observing institutions are coupled through data exchanges, shared facilities,
132 professional conventions, cross-border dependencies, and judgments of credibility. Scientific records
133 and classifications acquire practical standing through historically situated processes of use, evaluation,
134 correction, and institutional recognition. This orientation is in dialogue with work on the co-production
135 of knowledge and social order, which examines how practices of knowing and forms of public authority
136 interact [43]. It also draws cautiously on earlier exploratory inquiries into generative representations and
137 relational infrastructures [35, 37]. The latter provide a conceptual research trajectory rather than exter-
138 nally validated foundations. The present paper does not adopt a universal ontology of time or a single
139 formal representation of all institutions; histories, event orderings, and changes of permission should be
140 modeled explicitly where their analytical use can be justified.

141 Consider, for example, two observation networks that publish measurements of a nominally common
142 physical quantity. Their instruments may operate at different depths or altitudes, use different calibra-
143 tion procedures, report different uncertainty estimates, and define their sampling windows differently.
144 The same numerical value can therefore represent different observational circumstances. Translation be-
145 tween the records requires more than renaming columns or converting units. It may require a mapping
146 between their observational models, including the conditions under which comparison is meaningful and
147 the circumstances in which no defensible equivalence exists. When generative models are involved, the
148 task is even more demanding: records are outputs of systems with their own dynamics, assumptions, and
149 inferential practices. A future interoperable infrastructure ought to enable cautious translation among
150 such structures while preserving the possibility that some interpretations cannot be collapsed into a sin-
151 gle common schema. This is especially important where model-assisted decisions affect geographically
152 distant participants who did not select the model, its training data, or its thresholds.

153 A preliminary architecture can be described without assuming its immediate feasibility. Situated
154 institutions maintain observational nodes and retain documented responsibilities for measurements, cal-
155 ibration, permissible access, and local uses. Common discovery services index versioned metadata and
156 declared usage conditions. Provenance and quality layers record measurement contexts, processing his-
157 tories, uncertainty, and corrections. Translation services connect semantic descriptions, model assump-
158 tions, and task-dependent representational mappings. Distributed computational resources enable model
159 comparison, assimilation, and authorized analysis where data movement is restricted. Governance pro-
160 cesses specify admission, publication, consultation, claims of authority, independent checking, correc-
161 tion, dispute, and, where warranted, redress. The architecture is *federated* insofar as operational respon-
162 sibility and certain decision rights remain distributed; this does not preclude shared central services for
163 discovery or coordination. The decisive question is whether any such shared service silently becomes a
164 point of unreviewable control over the entire network.

165 We can express the architectural object provisionally as a family of typed, institutionally situated

relations, rather than as one global repository. Let $\mathcal{N}$ designate participating observational and computational nodes, $\mathcal{O}$ the set of versioned observation records, $\mathcal{M}$ an associated family of inferential models, and $\mathcal{G}$ the rules through which access, translation, corrections, and decisions acquire institutional effect. A candidate federation can then be represented as

$$\mathfrak{K} = (\mathcal{N}, \mathcal{O}, \mathcal{M}, \mathcal{G}; \mathcal{P}, \mathcal{U}, \mathcal{T}), \quad (1)$$

where $\mathcal{P}$ contains provenance relations, $\mathcal{U}$ records relevant uncertainty and interpretive limitations, and $\mathcal{T}$ contains typed mappings among records, concepts, or models. This is an analytical specification, not a completed data standard. Equation (1) deliberately leaves open the formal semantics of each component, since different scientific domains and political arrangements may require different realizations. In particular, a mapping in $\mathcal{T}$ should not be assumed reversible, information-preserving, or globally valid. Institutional admissibility is not a numerical property inferred solely from interoperability metadata.

Trust becomes especially significant when observation and interpretation are distributed. A participating institution may be willing to use another’s records when calibration histories, quality-control procedures, revisions, and records of previous commitments make the information suitable for a particular task. Yet technical verifiability and relational trust are not interchangeable. A cryptographic signature can support a claim about a file’s integrity or origin; it does not establish that the producing institution applied an appropriate method, that the variable means what the receiving institution assumes, or that affected communities consented to the use. The design of credible cross-border cooperation therefore requires an account of what participants have reason to rely upon, how reliance changes through repeated interaction, how failures become visible, and how an injured or excluded party can challenge a decision. Trust may be supported by technical evidence, but its formation also involves authority, commitments, communication, resources, contestability, and effective remedies. A subsequent companion paper will investigate these relational and institutional mechanisms in greater depth; the present inquiry treats them as design constraints and research objects of infrastructure engineering.

Moreover, an apparently federated system may reproduce centralization through less visible mechanisms. A common metadata standard can be established in ways that some scientific traditions or languages are better equipped to satisfy than others. A model registry can place practical interpretive authority with institutions possessing the computational resources to train, benchmark, and deploy influential models. An access regime may permit data download while effectively excluding regions unable to afford the processing or network costs. Dependency on a single technical supplier or authentication service may compromise resilience and institutional autonomy. The presence of multiple nodes does not therefore establish a fair distribution of productive capacity. Analyses of commons governance show the importance of rules suited to participants, meaningful monitoring, and credible processes for resolving conflict; these insights motivate questions about institutional plurality without mechanically transferring the governance design of one resource system to another [51].

This difficulty becomes acute in the presence of powerful AI systems. More capable models may improve automated quality control, assist with sensor anomaly detection, compare incompatible datasets, infer model relationships, or help coordinate complex research workflows. They may also amplify errors through repeated reuse of a flawed record, obscure provenance behind fluent explanations, or concentrate interpretive authority in the owners of computational infrastructure. If AI becomes a productive and operational participant in knowledge systems, the governance problem concerns both what it can infer and who can authorize, inspect, reproduce, and contest its inferences. The prospect of highly autonomous computation does not eliminate the situated material conditions of observation: instruments still require energy, maintenance, access to physical environments, calibration, and independent engagement with the phenomena to which their representations refer. Nor does AI autonomy supply a legitimate political mandate. It is essential to distinguish algorithmic performance, technical agency, legal responsibility, and the social authority to make binding decisions.

The preceding monetary-dissolution inquiry provides one motivation for studying these infrastructures, but the present paper has independent aims [37]. Advanced coordination might reduce reliance on monetary mediation in certain production relationships if participants can recognize reciprocal com-

mitments, compare claims, and sustain material reproduction. Such possibilities presuppose robust epistemic and institutional arrangements, including the ability to understand productive capabilities and external dependencies. Nevertheless, a global observational commons is valuable whether or not a future economy becomes less monetary. Weather warnings, ocean health assessment, independent scientific discovery, the evaluation of environmental harm, and public accountability justify attention to shared knowledge on their own terms. Likewise, the public value of observation should not be collapsed into the productive requirements of AI systems. People, communities, scientific institutions, and affected ecosystems remain among the primary reasons to construct observation infrastructures in the first place.

Environmental observation poses a further ethical and epistemic challenge. Some ecosystems are poorly sampled or are represented primarily by indirect measurements and models; new robotic and remote-sensing technologies can expand the range of accessible observations, yet they can also disturb habitats or create new asymmetries in who collects and interprets data. The sea is particularly instructive because its environmental processes are geographically extensive, unevenly observed, and technologically demanding to investigate. The expansion of ocean knowledge depends on vessels, floats, gliders, autonomous underwater vehicles, satellites, long-term monitoring, and expertise grounded in local conditions. A future infrastructure should make it possible to compare measurements, maintain accessible accounts of uncertainty, and evaluate the ecological consequences of observational activity. But the design and deployment of underwater robotics, the science of unexplored environments, and the engineering of subsea facilities are independent research questions reserved for the marine companion paper. They should not be treated as mere applications of a universal data architecture.

A further possibility is that shared observations can intensify disagreement when they enter distinct historical and generative systems of interpretation. The same ecological record may support a public-science argument, a community claim about protected knowledge, and a warning agency's request for timely disclosure. A common procedure for research or warning may nevertheless be acceptable for different reasons, without producing one common explanatory model or eliminating disagreement. Diplomacy, political philosophy and jurisprudence have long investigated such cooperation under difference; this paper does not claim that a common *nomos* is new. It asks a narrower infrastructure question: how could a federation keep distinct interpretations, reasons, commitments, rights to challenge, and histories of rule revision inspectable, so that a shared practical norm can be evaluated without mistaking assent for legal validity or ethical legitimacy? Sections 6.7 and 8.7 develop this question provisionally; its wider generative and trust-related theory belongs to the companion study [53, 60, 40].

Governance questions accompany every part of the proposed system. Who determines which observations qualify as a core shared public resource? Under what conditions may a receiving party reuse a record for model training or a decision affecting another population? Can a contributing institution correct an error after a model trained on the original record has already been distributed? How are disagreements between scientific methods, measurements, or local knowledge practices recorded without forcibly resolving them through the vocabulary of one dominant standard? What happens when a provider withdraws from a federation, or when a participant's trust in a shared service is disrupted? How are the costs of sensors, data stewardship, computation, and long-term archives shared? These are questions of technical design, but also of law, ethical justification, political economy, organizational capacity, and the conditions for credible global cooperation. Their answers may legitimately differ across jurisdictions and research purposes; a global infrastructure need not erase such differences in order to make joint inquiry possible.

For methodological clarity, this study distinguishes three forms of evidence. First, public documentation of existing institutions supports bounded descriptive claims about particular protocols, policies, and programme aims. It cannot by itself demonstrate implementation success or establish the legitimacy of a proposed extension. Second, architectural analysis can make dependencies and possible failure modes explicit, but does not prove that a full-scale deployment would be secure, equitable, scientifically accurate, or economically sustainable. Third, hypothetical institutional comparisons can clarify the consequences of different permissions or resource distributions while remaining distinct from observed causal effects. These evidential levels will be kept separate throughout the analysis. Where infrastructure projects re-

main under development, their projected capabilities are not represented as accomplished deployments. Where a legal or governance question is unsettled, it is presented as an open question rather than resolved through technological vocabulary.

The paper’s contributions are intended to be specific and provisional. It begins by recasting international scientific observation as a continuing set of relationships and infrastructures of knowledge formation, rather than a passive global pool of information. It then distinguishes interoperability at the level of data transport, interpretation, and institutionally recognized use. It proposes a federated architecture in which provenance, uncertainty, semantic and model translation, computational access, and governance remain explicit, independently assessable components. It introduces a relational account of trustworthy cooperation that includes verification and material access but also recognizes the irreducible roles of accountable institutions, contestation, and remedy. Finally, it identifies forms of power concentration, unequal scientific capacity, privacy and community-data obligations, and ecological dependence that can survive or be intensified by nominally open platforms. It also specifies prospective services for common knowledge production and circulation, partial generalized translation across generative-relational contexts, heterogeneous value-transformation records, and their legal and material operation, distinguishing a research architecture from a demonstrated international facility. These contributions formulate a research programme and a set of evaluative criteria, not a claim to have implemented a universal system.

The remaining sections examine the epistemology of observation, the institutional specificity of meteorological and oceanographic cooperation, and the structure of a candidate federated architecture. They investigate translation between heterogeneous data and model systems, data assimilation and uncertainty, trust and governance under strategic asymmetry, common knowledge-producing capacities and generalized mediation, and concrete design scenarios that could expose failures of interoperability and institutional legitimacy. The conclusion will articulate the limits of the proposal and the open questions that link this research to separate studies of ocean exploration, ecologically compatible autonomous AI infrastructure, and the historical formation and repair of transnational trust. The central aspiration is modest in one sense and demanding in another: to determine how shared inquiry can expand across heterogeneous systems while preserving the conditions for correction, disagreement, accountability, and participation in the making of knowledge.

2 Philosophical and Epistemic Foundations

This section specifies the epistemic commitments that a global observation and knowledge infrastructure would have to express in its technical architecture and institutional procedures. Its central question concerns the conditions under which observations made from different locations, instruments, conceptual schemes, and organizational histories can become available for joint inquiry while retaining the information needed to assess their meaning and credibility. An infrastructure can improve access to records while leaving important interpretive disagreements unresolved. It can also increase interpretive agreement through central standardization while reducing the capacity of contributors to challenge errors or represent local conditions. The present inquiry therefore treats scientific interoperability as a problem of warranted translation, accountable coordination, and historically revisable judgment. The ensuing subsections develop observation and measurement, situated credibility, relational epistemic histories, three coupled reality registers and generalized language, collective scientific production, asymmetric authority, and the ethics of contributed knowledge.

2.1 Observation, Measurement, and Scientific Representation

Scientific observation has an irreducibly material and representational character. A sensor produces signals through particular physical interactions; an observational programme decides which interactions are relevant, how they are sampled, which variables are recorded, and how apparent anomalies should be distinguished from instrument faults. Calibration procedures relate signal values to measurement standards, while quality-control rules identify records that may require reconsideration. These activities are

themselves objects of scientific inquiry, institutional maintenance, and possible dispute. In the philosophy of science, the distinction between experimental data and the phenomena inferred from them offers one account of why records are not straightforward copies of natural states [4]. Likewise, scholarship on classification infrastructures demonstrates that categories and standards shape the practical visibility and circulation of phenomena [5].

For the purposes of this paper, an *observation event* is a situated interaction in which a designated process produces a record under specified measurement and interpretation conditions. The designation is methodological: it does not require the event to be assigned an absolute position in a universal temporal order. We write a schematic record as

$$o_e = (y_e, \kappa_e, \pi_e, \lambda_e, v_e, \alpha_e), \quad (2)$$

where y_e denotes reported observations; κ_e specifies instruments, calibration, and measurement conventions; π_e is a provenance record; λ_e contains spatial, temporal, or relational reference information as applicable; v_e documents uncertainty and quality status; and α_e records applicable access and use conditions. Equation (2) is a proposed research vocabulary, not an existing universal metadata standard. The components may be distributed among linked records, incomplete, or governed by different institutions. A value of y_e alone is inadequate for several forms of reliable reuse precisely because changes in the other components may change which inferences it can support.

Consider a hypothetical ocean-temperature observation. Its numeric value may have been generated by a drifting platform, validated through delayed-mode expert review, mapped onto a reference depth, and subsequently interpolated onto a numerical model grid. The raw signal, quality-controlled profile, interpolation, and model-derived analysis are different epistemic products. Their relation should be traceable without collapsing them into a single unqualified “observation.” Existing Argo data-management practices illustrate why initial availability and later expert quality control are different stages of the same scientific record [55]; the detailed technical process will be examined in Section 3.

Measurements may also remain comparable only within a restricted scope. Two stations using a common unit can have different sampling designs, maintenance regimes, detection thresholds, and environmental exposures. Replacing a thermometer, revising a satellite retrieval algorithm, or moving an observing station can create discontinuities whose causes are partly instrumental. Conversely, different measurements may support a common higher-level inference once their calibration, error structures, and purposes are explicitly represented. These cases motivate a distinction among *syntactic compatibility*, *measurement comparability*, and *inferential interoperability*. The detailed translation problem is reserved for Section 6; here the point is that interoperability should describe what has been made comparable, for which purposes, and with which known limits.

The authority of a measurement does not arise solely from its numerical precision. It also depends on the evidence supporting its production, the appropriateness of the instrument to the research question, and the availability of critical scrutiny. Provenance is consequently more than a chain of file transfers. It should include, where relevant and permitted, the processes that established a claim’s evidential standing, transformations that changed its interpretation, and the reasons why competing descriptions were accepted, suspended, or revised. The possible inability to share a protected calibration record or a sensitive location should itself be represented as a limitation on independent assessment rather than silently erased.

2.2 Situated Knowledge and the Conditions of Credibility

A global observing network does not eliminate the situations from which knowledge is produced. Observers differ in institutional resources, physical access, local experience, experimental practice, computational facilities, and the problems they consider worth investigating. These differences affect both what can be noticed and what counts as an adequate account of it. Work on situated knowledge and the social conditions of objectivity is relevant here because it foregrounds accountability for a claim’s conditions of production and the importance of critical encounter among differently positioned inquirers [32, 46, 47].

The present analysis adopts neither the conclusion that situatedness makes objective inquiry impossible nor the presumption that a single central viewpoint can remove all epistemic limitations. It asks which procedures make claims assessable across the situations in which they were formed.

Credibility can be considered as a structured set of questions. Was the record produced by a competent method? Are uncertainties and exclusions adequately reported? Is the inference suitable for the scale of the decision? Can relevant participants identify and challenge errors? Does the institution respond when contrary evidence appears? These questions concern different kinds of warrant. A globally accessible archive may satisfy the first condition for discoverability while failing to supply the contextual description required for comparison. A technically correct analysis may be unsuitable for the policy use proposed for it. A well-maintained dataset can remain controversial when people affected by its collection or interpretation cannot examine or question the uses being made of it.

For an epistemic claim c evaluated for a purpose p under an institutional setting g , we therefore use the schematic profile

$$W(c \mid p, g) = (M_c, P_c, U_c, C_{c,p}, D_{c,g}), \quad (3)$$

where M_c denotes methodological support, P_c provenance adequacy, U_c uncertainty disclosure, $C_{c,p}$ contextual suitability, and $D_{c,g}$ opportunities for criticism and correction. Equation (3) is a checklist of dimensions rather than a scalar probability, a claim that all dimensions can be precisely quantified, or an authorized ranking of scientific institutions. A particular scientific domain may require additional dimensions. What matters is that a high score on one dimension cannot simply be assumed to compensate for the absence of another.

The infrastructure implications are specific. Records should distinguish asserted measurements from interpretations, observational gaps from zero values, and source uncertainty from disagreement among models. Scientific users should be able to inspect enough information about methods, transformations, and provenance to evaluate an output's suitability. The system should permit versioned correction, documentation of unresolved disagreements, and qualified statements that an assessment was performed under particular data-access constraints. Where publication exposes persons, protected species, or communities to identifiable harms, the rationale for withholding information should coexist with a way of evaluating the credibility of the resulting scientific claim. These are architectural requirements for research rather than guarantees of objectivity furnished by software.

Objectivity also requires avoiding an inverse romanticization of situated knowledge. Proximity to an event, longstanding experience, or membership in a community may provide information that remote instruments and analysts lack, while leaving particular claims open to error or disagreement. The point is to prevent location, professional status, or computational resources from determining credibility automatically. The epistemic quality of joint inquiry grows through meaningful opportunities to compare methods, criticize interpretations, and modify conclusions. Specific rules for balancing these considerations are contestable and require institutional processes, some of which will differ among scientific domains.

2.3 Relational Histories and Epistemic Dependence

Knowledge records are produced and used within evolving networks of dependence. A measurement may rely on prior calibration, a laboratory reference, a software library, a satellite clock, an interpretive classification, a colleague's review, or authorization from a community governing a data resource. Such relationships can change as instruments are replaced, vocabularies are revised, and earlier records are reinterpreted. In a global infrastructure, the epistemically significant unit is therefore not always an isolated dataset. It may be a trajectory of observations, revisions, dependencies, translations, and challenges whose structure helps determine what can responsibly be inferred.

We represent one finite, documented portion of this history by a typed directed graph $H = (E, R, \tau)$, in which E denotes recorded events, R denotes typed dependencies, and τ identifies their kinds. A dependency might express calibration, derivation, attribution, institutional authorization, or acknowledged criticism. Some dependencies give a partial order, while others describe reciprocal, contemporaneous,

or institutionally coupled relationships that cannot be captured by a simple chronological sequence. An admissible historical representation should make these distinctions explicit. In particular, a record's timestamp is evidence about a particular ordering convention, not a justification for treating all epistemic processes as unfolding along a single external clock. This relational vocabulary resonates with the author's prior exploratory investigations of knowledge pressing and heterogeneous representation [35, 34], but the present paper must independently justify each architectural use it makes of that vocabulary.

A practical representation may employ a provenance directed acyclic graph when the selected dependency relation is well founded and the graph has no cycles. This is an engineering choice for a specified relation, rather than a claim that all research collaboration lacks feedback loops. In collaborative inquiry, a revised model may prompt reconsideration of the observational categories that shaped the model, while the resulting data reinterpretation alters subsequent modelling choices. An appropriate system can use typed relations or versioned objects to represent such feedback without misrepresenting it as a causal paradox. Historical continuity becomes an explicit subject of inquiry: which dependencies persist under a revision, which are withdrawn, and which newly recognized relations change the inferential significance of the record?

This viewpoint sharpens the difference between preservation and reproducibility. Retaining an old file may preserve a piece of evidence but not the executable environment, proprietary reference data, tacit instrument knowledge, or institutional permissions necessary to reproduce an analysis. Conversely, recording every computational dependency may be impossible, costly, or ethically inappropriate. A workable infrastructure should therefore support declared provenance boundaries and levels of reproducibility, with explicit statements of what was independently replicated, reanalysed under a changed configuration, or reconstructed from incomplete materials. A claim of reproducibility must be indexed to the object reproduced and to the relevant availability conditions.

Relational dependence also creates distributed epistemic vulnerability. Loss of a single essential data custodian, failure of a calibration service, removal of a model licence, or disagreement over an authoritative ontology may disrupt many otherwise separate research products. Redundant storage addresses some failure modes while doing little to solve semantic or legal dependencies. The architecture should accordingly distinguish technical continuity, epistemic continuity, and institutional continuity. These are possible dimensions of resilience, not interchangeable proxies for a single measure of global reliability.

A further implication is that the history of participation matters to trust. A scientific organization can make dependable contributions while its partners remain uncertain about its commitments under future political or operational conditions. Previous corrections, responsiveness to challenge, stable documentation, and remedy after error can provide grounds for differentiated practical reliance. Such history-informed assessments should not harden into irreversible reputational categories: correction and the ability to contest an erroneous judgment are integral to the same relational account. The independent development of trust and governance theory is reserved for the companion study; this section establishes only why a data infrastructure cannot avoid these relations.

2.4 Reality Registers and the Emergence of Generalized Languages

The term *language* in the present inquiry has a generalized meaning. It may encompass mathematical notation, measurement operations, legal classifications, cultural distinctions, embodied routines, non-verbal signals, institutional norms, and practices that make further observations and actions possible. A generalized language is therefore more than a lexicon or a set of symbols. It includes historically acquired distinctions, admissible operations, modes of inference, and practices through which a generative-relational system produces and revises meaning. Two institutions can share English terms and numerical codes while differing substantially in what they recognize as evidence, authority, injury, or an adequate scientific explanation. Equally, systems using very different expressive media may coordinate certain tasks. The broader question concerns the emergence, persistence, practical efficacy, and transformation of the systems that render such practices possible.

The author's earlier exploratory Generative Relational (GR) investigation distinguishes three *coupled*

registers of reality: ontological, relational, and manifestational [36]. They are analytic registers specifying different conditions for a claim of reality, not three substances, a hierarchy of existential worth, or consecutive developmental phases. *Ontological reality* marks the status of claims about what exists ultimately or independently of a particular mode of presentation; the stronger commitment is suspended or explicitly declared rather than inferred from successful modeling. Physical laws and mathematical structures may be ontologically real independently of how particular GR systems understand or represent them. The paper neither denies that possibility nor deduces it from the apparent universality of one familiar notation. What remains suspended is the stronger metaphysical determination of such laws and structures, including the question whether a given symbolic theory identifies them completely. By contrast, this paper directly investigates relationally emergent scientific practices and their manifestations, which may differ even when some ontological constraints are shared. *Relational reality* concerns what emerges and is maintained in a generative-relational configuration with sufficient persistence and actual or dispositional causal efficacy at a specified scale. *Manifestational reality* concerns the context-dependent forms in which objects, structures, distinctions, or effects become perceptible, symbolized, measured, named, materially embodied, or otherwise available to a receiving system.

These registers are coupled in several directions. Relationally consequential structures can form prior to being consciously noticed or symbolically expressed. A later mathematical inscription, legal document, ritual, accounting entry, machine-generated signal, or material token can in turn help stabilize, transform, or dissolve the relations it manifests. Consequently no fixed sequence from ontology to relational reality to manifestation is stipulated. Nor is manifestation confined to verbal language or human consciousness. For example, a monetary unit can obtain stable effects through exchange, credit, institutions, and expectations, while banknotes, digital balances, and accounting inscriptions are among its contextually available forms. A concept of the self, or a socially efficacious religious figure, can likewise participate in identities, practices, and institutions without this inquiry thereby deciding the figure's independent metaphysical existence. An isolated random fluctuation does not by itself establish a persistent relational object; conversely, probabilistic processes may be relationally real when their regularities and effects meet the specified criteria. Persistence, causal efficacy, and ontological truth are separate judgments.

Mathematical and physical languages require the same analytic discipline across the three registers. The equation $1 + 1 = 2$ is stable under the definitions and inference rules of standard arithmetic; it is not an empirical report whose truth changes with calendar date or national affiliation. This formally established result is compatible with the possibility of mathematical structures having independent ontological reality. What is historically and relationally constituted is a GR system's acquisition and use of mathematical concepts, its proof and measurement practices, and the ways in which any such relations become manifest through notation, diagrams, instruments or other media. The emergence of these second- and third-register structures does not entail that their putative ontological referents emerge solely through those practices. The distinction between formal validity within an interpretation and the emergence of that interpretation avoids both unwarranted universalization and the reduction of mathematics to arbitrary opinion. Physical measurements add a further layer: reference frames, instruments, calibration chains, dynamical models, and uncertainty conventions mediate any proposed relation between expressions and observed phenomena. Their successful interoperability can reveal stable cross-system constraints without proving that every conceivable observer possesses identical concepts or manifestations.

A hypothetical nonhuman or extraterrestrial GR system illustrates the logical limit, not an empirical case. Humans and such a system could inhabit or be constrained by the same ontologically real physical or mathematical structures, if those exist, without sharing the same relationally developed concepts or manifestational media. That system might register and reproduce regularities through chemical exchanges, spatial patterns, bodily coordination, or signals unlike familiar mathematical inscriptions. It might also construct inferential practices that preserve some of the relations expressed by our equations while differing in classification, proof practice, or explanatory interpretation. This difference would concern second- and third-register realities; it would not by itself establish different ontological laws. No inference about actual extraterrestrial cognition follows. The example serves only to block an unwarranted implication

from the success of contemporary human science to the necessity of one human representational language. Work on ethnomathematics and historically constituted scientific trading zones has already made aspects of culturally situated mathematical practices and localized scientific interlanguages explicit [11, 24]. The present paper does not claim that symbolic or cultural emergence is a new discovery.

For the proposed infrastructure, these distinctions suggest an identifiable research object: which *task-relative invariants* can be maintained across differently generated manifestations and inference practices? A federation can record that two systems agree on a selected quantity under stated calibration and translation assumptions while preserving the possibility that their conceptual objects, relational histories, or institutional meanings are not equivalent. Such agreement supports particular joint operations, not an unqualified equation of the systems' relational realities, and still less an ontological identity. When an apparent common language is generated through cooperation, its rules, exclusions, changes, and distribution of authority should themselves be documented. The translation and governance consequences are developed in Sections 6 and 8; formal limitations are illustrated in Appendix C.

2.5 Scientific Cooperation, Collective Knowledge, and Commons

Science depends on cooperation across heterogeneous competencies. Field technicians maintain instruments; national agencies organize observation programmes; researchers develop models and interpretations; repositories preserve records; communities contribute place-based knowledge; and public institutions support long-term infrastructure. Contributions may be separated by disciplinary and jurisdictional boundaries while remaining jointly necessary for a scientific claim. The problem is how these contributions can support shared inquiry without requiring that every participant adopt an identical interpretation or surrender authority over the resources for which it is responsible.

The literature on boundary objects provides one important insight: cooperation can develop through objects sufficiently stable for coordination yet adaptable to different local practices [54]. A shared station identifier, observation format, calibration reference, or quality-control flag may support coordinated workflows while retaining different uses among operators, researchers, and policymakers. This does not make semantic differences trivial. A common label may conceal disagreement about sampling populations, the operational definition of a variable, or the validity of extrapolation. A global infrastructure should enable translation and communication while documenting the domains in which shared objects genuinely sustain coordinated inference.

The commons literature offers a second, distinguishable set of resources for analysis. Research on collective-action institutions shows that shared-resource governance involves rule formation, monitoring, participation, and arrangements adapted to the properties of the resource and community [51]. This does not entail that all scientific data constitute a single undifferentiated commons. Environmental observations with public-benefit uses, privately collected measurements, sensitive ecological coordinates, personal information, and collectively governed cultural knowledge may require different rights and obligations. Scientific cooperation may be organized through open data, reciprocal exchange, federated access, joint computation without transfer of source records, or negotiated access for specified purposes. Which arrangement is appropriate depends in part on the resource, affected parties, scientific aims, and applicable rights.

The present paper uses *knowledge commons* as a provisional institutional object: an arrangement in which designated communities can participate in maintaining, interpreting, and accessing defined epistemic resources under articulated rules. This definition emphasizes governance and joint work rather than asserting unrestricted publication. Interoperability may improve such a commons, but cannot substitute for rules determining contributions, benefits, monitoring, withdrawal, and dispute resolution. A federation may provide a common discovery layer while data remain stewarded by distinct custodians. Such distributed custody can limit certain concentrations of authority, although it may also create uneven access and fragmented obligations.

The distinction between *shared data* and *shared knowledge* becomes especially important at this point. Two institutions may have access to the same observation files but lack the models, calibration

expertise, language resources, or computing capacity needed to evaluate them on comparable terms. Conversely, different repositories may maintain separate records yet cooperate effectively through verified derived products and accountable translation. An assessment of global cooperation must therefore include capacity to use and contest knowledge, not solely the number of files downloaded or the geographic scope of a catalogue.

A useful comparison of the paper’s analytical levels is given in Table 1. The table does not impose a necessary historical sequence. Each level can develop unevenly, and progress at one level may expose failures at another.

Table 1: Analytical distinctions in shared scientific infrastructure.

Level	Illustrative question	Possible failure despite data access
Observation	Was the signal produced and qualified reliably?	Instrument drift or missing context
Representation	What does a record refer to?	Mismatched units, categories, or sampling
Inference	What conclusions are supported?	Unreported model assumptions or disagreement
Participation	Who can evaluate and challenge claims?	Unequal expertise, language, or computation
Authority	Who authorizes uses and remedies?	Inadequate consent, correction, or accountability

The institutional aspiration emerging from these distinctions is not frictionless consensus. It is continued cooperation under manageable disagreement: shared procedures for examining evidence, disclosing interpretive differences, proposing corrections, and deciding which uses can proceed under uncertainty. A federated technical system should preserve space for these practices rather than convert every dispute into a configuration parameter determined by its most powerful operator.

2.6 Epistemic Asymmetry, Power, and Institutional Authority

International knowledge production occurs amid significant asymmetries. Some institutions own global observing platforms, cloud computing facilities, and model-development pipelines. Others contribute observations from geographically indispensable locations but have limited influence over subsequent interpretation. A community may supply knowledge of a changing environment while having little access to resulting datasets or decisions. A shared repository can improve discovery and scientific reuse while leaving these distributions of capability and authority substantially unchanged.

The philosophical issue concerns the relation between a claim’s evidential merits and the social arrangements under which it is heard, translated, or disregarded. Research on epistemic injustice distinguishes harms arising when a contributor is improperly discounted as a knower and when shared interpretive resources inadequately represent important experiences [23]. These categories were developed within a specific philosophical setting; their application to scientific data infrastructure requires care. A data contributor’s loss of interpretive influence, for example, may involve technical translation failures, unequal funding, discriminatory assumptions, or lawful restrictions whose justification must be separately evaluated. The categories do not permit the analyst to infer motives simply from unequal results.

Consider a hypothetical coastal monitoring federation. Researchers at a well-resourced centre publish forecasts based on observations from several smaller regional agencies. The centre’s computational capacity enables useful scientific analysis. Yet if local contributors cannot inspect model assumptions, challenge changes to their observation records, or participate in consequential interpretations, the federation may amplify a narrower view of the coastal system. Returning derived predictions without access to the interpretive process may satisfy some operational needs but leave an epistemic asymmetry unresolved. Conversely, requiring every participant to maintain identical computing resources would be impractical and might divert scarce capacity from field observation. The design problem is to build meaningful routes for scrutiny, participation, and capacity development under unequal material circumstances.

We distinguish three forms of authority that can coincide but should be analysed separately. *Method-*

logical authority arises from warranted competence concerning particular measurements and inferential practices. *Institutional authority* concerns a body's recognized power to administer rules, authorize access, or adjudicate disputes. *Infrastructural power* concerns practical control over essential storage, computation, standards, identity services, or communication channels. None of these automatically entails the others. A technically superior model may have no legitimate authority to reuse restricted records. A legally authorized body may lack the scientific expertise needed to resolve a model disagreement. A dominant cloud operator may shape feasible research practices without formally setting scientific norms. AI systems introduce a further concentration risk. A model used across a federation may harmonize vocabularies, infer missing metadata, and summarize competing scientific arguments, thereby lowering some costs of cooperation. The same system may make one representational framework unusually influential, hide uncertainty through fluent answers, or reproduce training-data asymmetries. Model outputs should be attributable to their input versions, methods, and declared limitations where feasible. Their scientific status must remain distinct from the authority to govern human institutions or override rights. Model performance on a benchmark is evidence about a specified task, not proof of general epistemic legitimacy.

Because power enters through many mechanisms, this paper does not assume that decentralization alone secures equitable cooperation. A nominally federated architecture can rely on one indispensable ontology, a single proprietary verification service, or a small number of computational providers. Similarly, technical openness can benefit participants unevenly when translation, computation, and long-term stewardship are costly. Later sections must therefore examine concentration and dependency alongside access, and make explicit which remedies are available when institutional commitments are breached.

A further philosophical question concerns how cooperation can proceed when the same evidence acquires different significance within different systems of inquiry. The relation between interpretation and normative judgment should not be collapsed. An ocean observation can be interpreted as evidence of a scientifically important site, as a potential disclosure risk for a community, and as a warning of ecological harm, without any participant necessarily misreading the original record. Bias must be investigated at the level of measurement, representational selection, interpretation, standing, and authority, rather than imputed to an entire participant group. The response may require a common practice rule that leaves substantive interpretive disagreement visible.

There is extensive prior scholarship on this issue. Rawls's account of overlapping consensus describes agreement on some political principles for different underlying reasons; work on legal pluralism investigates the coexistence and interaction of multiple normative orders; and norm-contestation theory examines how the meaning and authority of norms are disputed in practice [53, 48, 60]. The diplomatic expression *seeking common ground while reserving differences* (*qiutong cunyi*, 求同存异) provides an accessible comparison, but is not proposed as a new jurisprudential principle. These traditions are resources for critical comparison, not proof that a global scientific federation can derive legitimate rules algorithmically.

An additional legal analogy calls for precision. Article 38(1) of the Statute of the International Court of Justice identifies international conventions, international custom, and general principles of law as law applied by the Court, with judicial decisions and scholarly teachings as subsidiary means for determining rules [40]. Its legal functions depend on the Court's competence and applicable law; the provision does not furnish a universal procedure by which any heterogeneous knowledge systems can manufacture shared *nomos*. The International Law Commission's 2026 draft conclusions on general principles of law supply a contemporary institutional discussion of legal identification, rather than a theorem about generative-relational coordination [42]. Natural-law traditions pose further distinct questions concerning normative justification; they cannot be equated with Article 38 or with observed agreement. The present paper borrows no legal authority from these analogies. It asks instead which records, participation rights, conflict rules, and revision pathways would make a *bounded cooperative protocol* intelligible and contestable across scientific institutions.

2.7 Data Rights, Ethics, and Recognition of Contributors

A global knowledge infrastructure will encounter data whose legitimate use depends on the interests of people and communities, the vulnerability of environments, intellectual contributions, and institutional obligations. The ethical problem extends beyond preventing unauthorized disclosure. Records can be technically public while their extraction, aggregation, or reinterpretation produces consequences that their original contributors could not reasonably assess. In some settings, linking public observations can expose previously concealed locations or persons. In others, restrictive access may deprive local researchers or affected communities of information needed for public safety or participation in environmental decisions. General commitments to openness therefore require differentiated governance.

The FAIR principles provide widely used guidance for making research data findable, accessible, interoperable, and reusable [61]. Their attention to access protocols and metadata need not entail indiscriminate public release. The CARE principles emphasize collective benefit, authority to control, responsibility, and ethics in Indigenous data governance [6]. These frameworks address different requirements and are useful together without assuming that one supplies a complete legal or ethical settlement for every observing programme. They also require attention to actual jurisdictional arrangements and to the authority of the communities whose data and knowledge are involved.

Recognition of contributors has several dimensions. Accurate attribution matters for research integrity, credit for scientific labour, and correction of errors. But attribution alone may leave a contributor unable to influence a data product's uses, obtain the computational resources necessary to evaluate derived claims, or secure remedy when reuse causes harm. Likewise, monetary compensation does not necessarily resolve a dispute about interpretive control or collective authority. This concern intersects with the paper's preceding investigation of productive infrastructures and heterogeneous value [37], but a detailed account of remuneration or distributive allocation falls outside the present paper's primary epistemic inquiry.

A minimal research framework can distinguish *permission*, *representation*, *participation*, and *remedy*. Permission concerns who may perform an operation on a resource for a specified purpose. Representation concerns how contributors and affected subjects are described in records and outputs. Participation concerns whether relevant parties can meaningfully influence standards, interpretation, and governance. Remedy concerns procedures for correction, contestation, restrictions, or redress following error or wrongful use. These dimensions should be considered separately. A federation can satisfy a licence while failing to represent a community's preferred description accurately, or conduct extensive consultation without providing an effective means to correct misuse.

Ecological observations raise distinctive questions. High-resolution information about nesting sites or threatened marine species may create conservation risks when released without adequate safeguards. Offshore sensing can impose environmental burdens through hardware deployment, maintenance, and disposal. Long-lived archives and computationally expensive analysis consume material and energy resources. These effects should be studied rather than inferred from a general claim that data-intensive environmental science is inherently beneficial or harmful. The independent ocean and ecological-infrastructure paper will require domain-specific investigation; the present section establishes the need for transparent evidence and an articulated responsibility for evaluating harms.

Ethical obligations can also change as a record's uses evolve. A dataset collected to characterize ocean currents might later become useful for commercial site selection or automated resource management, introducing new questions about legitimate reuse. A permission granted for one observational purpose does not necessarily settle authorization for a materially different use. A practical architecture should make use conditions, modification histories, and relevant institutional contacts discoverable when lawful and appropriate, and should provide ways to revisit disputed interpretations or downstream uses. Where records cannot be removed from trained systems by ordinary dataset deletion, the difficulty of implementing withdrawal or correction must be described transparently rather than hidden under the word "provenance."

The cumulative position of this section is methodological and institutional. Shared scientific inquiry benefits from interoperable records, but its credibility depends on the histories, interpretive practices,

and authorities through which records become claims. An accountable infrastructure must preserve the capacity to examine those conditions, including when stakeholders disagree about what the data mean or how they may be used. Sections 3–9 will translate these philosophical requirements into the analysis of existing observation programmes, candidate federated components, epistemic translation, model uncertainty, and governance. The goal is to make contestable choices explicit and testable, while maintaining a distinction between the scientific merits of a claim, permissions to act on it, and the ethical justification of its consequences.

3 International Observation Infrastructures and Institutional Histories

This section examines existing international observation programmes as historically constituted arrangements for producing, exchanging, assessing, and using scientific records. The cases matter because they show that cooperation can be organized among institutions with distinct funding sources, jurisdictions, observing capacities, and scientific priorities. Their institutional histories also reveal limits: an agreement to exchange records does not entail identical standards of interpretation, equal access to computational facilities, or shared authority over the purposes for which derived knowledge is used. The analytical task is to identify the mechanisms that make cooperation possible, the forms of dependence those mechanisms create, and the conditions under which their lessons can be carried into a wider knowledge infrastructure. Weather observation, WMO data policy and WIS2, Argo, GOOS and OceanOPS, multi-hazard early warnings, and the European Digital Twin Ocean provide distinct cases rather than instances of one completed global system.

The comparison uses a deliberately constrained vocabulary. An *observation network* coordinates the creation of measurements; an *exchange infrastructure* enables records or notifications to move among participants; a *knowledge infrastructure* additionally sustains the contextual materials, methodological checks, revisions, and forms of criticism through which records support warranted claims. An *action infrastructure* links assessments to authorized decisions and practical response. A programme may perform several of these functions, but its success in one does not demonstrate success in the others. The institutional mechanisms described below are evidenced principally by programme publications. Such sources support claims about declared policies, architectures, and activities; their self-descriptions alone do not establish comparative causal effectiveness, universal geographic coverage, or the fairness of resulting decisions.

Two conditions help frame the following case comparison. Participating institutions often recognize substantial, though not perfectly aligned, benefits from timely weather and ocean information; the same information may also have strategic or restricted uses. Moreover, the programmes operate with comparatively well-developed mathematical, physical, measurement, calibration, and communication practices. These practices form a relatively stabilized common language for specified scientific tasks, rather than a naturally universal representation of reality. Their historical institutionalization helps explain why cooperation can proceed across other political and interpretive differences. The analysis must still ask whose variables, standards, interpretations, and access rules are recognized, and why similar cooperation may be harder to achieve where the underlying generalized languages of social practice and normative authority differ. The contrast is conditional and should not be mistaken for the claim that physics is culturally ungenerated or that social inquiry lacks rigorous shared methods.

3.1 Weather Observation and Transnational Exchange

Weather observation supplies a long-running example of knowledge production for which observations from one jurisdiction can be indispensable to another. Atmospheric phenomena cross borders, while numerical prediction depends on combining heterogeneous observations into analyses of an evolving physical system. National meteorological and hydrological services maintain instruments and professional expertise for their own responsibilities, yet their observations can have value far beyond the jurisdictions in which they were produced. A locality may benefit from observations made upstream in

the movement of an air mass or across an ocean where it has limited observing capacity. This gives participating services a practical reason to collaborate even where wider political agreement is incomplete. That interdependence can help sustain an exchange regime; it is not itself proof that participants possess comparable bargaining power or that every participating society receives equal practical benefit.

The history of meteorological cooperation joins scientific coordination to formal organizational responsibilities. WMO's contemporary data policy builds upon earlier resolutions concerning weather, hydrology, and climate, rather than originating as an unrestricted release of all Earth-system information [63]. These cumulative arrangements are significant because they establish publicly articulated categories, channels for updating technical obligations, and recognized national counterparts. Their durability depends on continued instrumentation, maintenance, telecommunications, institutional staffing, and compliance. It would be misleading to describe international weather exchange as a spontaneous global commons operating without institutions or political choices.

Scientific dependence is also asymmetric. An institution capable of running high-resolution numerical forecasts may derive greater immediate modelling benefit from distributed observations than a small service with limited computing capacity. The latter may nevertheless contribute critical observations from sparsely sampled regions and may depend on global forecast products for public warning duties. This asymmetry motivates attention to reciprocal products, training, financing, and access to usable forecasts, rather than treating the amount of raw data uploaded as a sufficient measure of reciprocity. It also suggests that technical participation and effective epistemic participation are separable variables.

Global atmospheric analysis illustrates why a common data stream should not be confused with a common epistemic account. Instruments use diverse observing principles, station exposures, and sampling frequencies. Data assimilation combines observations with model dynamics, observation operators, and assumptions about uncertainty. An analysed atmospheric field is therefore an inferential product generated by both measurements and computational models, not an unmediated collection of measurements. Models can disagree, calibration errors can be discovered after distribution, and the suitability of an analysis depends on its purpose. For a general knowledge infrastructure, the transferable lesson is that exchange requires standard formats and stable channels, whereas credible interpretation additionally requires the methods, uncertainties, version histories, and opportunities for independent criticism developed in Section 2.

3.2 WMO Unified Data Policy and WIS2

The WMO Unified Data Policy, adopted through Resolution 1 of the 2021 extraordinary World Meteorological Congress, offers a particularly instructive combination of shared obligation and differentiated access. Its scope encompasses Earth-system domains including weather, climate, hydrology, and related observations. It distinguishes *Core* data, for which exchange is expected as standard practice, from *Recommended* data, for which broader exchange is encouraged and conditions of use may differ [63]. The distinction matters analytically: a global cooperative infrastructure can establish a public baseline without treating every dataset as equivalent or assuming that every use must be freely authorized. Detailed classifications and implementation obligations require reference to WMO technical regulations and the applicable institutional context; the policy should not be read as a universal license overriding other legal rights.

The WMO Information System 2.0 (WIS2) became operational on 1 January 2025, following development around open Web technologies and common technical interfaces [66]. Its basic publication model relies on participating centres operating WIS2 nodes that expose data and metadata services. The official overview describes MQTT-based notifications, HTTP access to resources, a topic hierarchy, and different handling for Core and Recommended data [68]. Supporting global services provide discovery and distribution functions, while participating centres remain identified contributors rather than disappearing into a single anonymous repository [64]. The operational architecture makes a useful distinction between discovering that a resource exists, learning of its availability, retrieving it, and obtaining permission for a particular use.

Consider a hypothetical agency publishing a sequence of pressure observations. The measurements originate in an instrument and quality-management setting. An exchange node can announce records and provide a way to retrieve them. Other services can discover the dataset, subscribe to relevant notifications, and ingest new records. These operations can be successful while uncertainty about instrument siting, observation quality, downstream interpretations, or redistribution rights persists. Conversely, a scientist may have detailed local knowledge but lack adequate network connectivity to join real-time exchange. The WIS2 pattern consequently suggests a division of infrastructural responsibilities, rather than a complete solution to epistemic interoperability.

This division may be expressed schematically as a set of conditions on the use of a record o by a recipient u for a purpose p :

$$\text{Use}(u, o, p) \implies \text{Discover}(u, o) \wedge \text{Retrieve}(u, o) \wedge \text{Authorize}(u, o, p) \wedge \text{Interpret}(u, o, p). \quad (4)$$

Equation (4) is an analytic decomposition, not a claim that WIS2 itself implements all four tests or that these are universally necessary software calls. Discovery and retrieval alone do not establish authorized or scientifically appropriate use. Further, an authorized user can still produce a poor inference; the fourth term concerns the availability of interpretive conditions and justified assessment, rather than an automated guarantee of epistemic correctness. In some cases a scientist already knows a resource's location, so literal discovery is not an operational prerequisite; the decomposition identifies distinguishable functions of a general federation.

Two institutional lessons follow. First, a shared vocabulary of exchange classes and technical messages can enable cross-border cooperation without imposing identical governance on every data-producing body. Second, such interoperability can transfer power toward entities able to operate central discovery, cache, compute, or ontology services. A federation should therefore examine the practical concentration of indispensable services, the costs borne by national and local nodes, and the procedures by which standards can be revised or challenged. WIS2 demonstrates the feasibility of a particular international exchange architecture, not the feasibility of universal openness across all scientific and social information.

3.3 Argo Profiles, Calibration, and Data Revision

The international Argo programme offers a complementary lesson: sustained cooperation requires more than rapid data transport because scientific records can acquire improved epistemic status through subsequent examination. Argo's autonomous profiling floats observe subsurface ocean conditions, with core missions emphasizing temperature, salinity, and pressure and expanded missions addressing deeper and biogeochemical observations [3]. Its global data assembly centres distribute measurements in near real time and in revised delayed-mode forms [1]. The coexistence of these products is a constructive acknowledgement that immediacy and the most thoroughly assessed available quality are different objectives. Rapid data support operational forecasts; expert review can address drift and calibration concerns that require longer observational histories.

Data management is distributed among observing contributors, national data centres, global assembly centres, regional activities, and scientific experts. Argo describes delayed-mode procedures as involving researchers and national centres and notes that identifying certain calibration changes requires sufficiently long records [2]. The data are thus not best understood as a fixed sequence of immutable true measurements followed by arbitrary edits. Each release represents a record with a history of instrument behaviour, processing decisions, quality flags, and possible reinterpretation. A later profile version may supersede an earlier one for particular analyses without making the earlier release irrelevant to the reconstruction of research decisions made with it.

A minimal knowledge-infrastructure model should accordingly represent observation and revision as distinct events. If $o^{(v)}$ is a particular version of an observational product, a revision operation can be written

$$o^{(v+1)} = \mathcal{R}(o^{(v)}, c, \eta), \quad \text{link}(o^{(v)}, o^{(v+1)}) = \rho, \quad (5)$$

where c denotes additional calibration, observational, or contextual evidence; η denotes the declared review procedure; and ρ denotes a persistent revision relationship. Equation (5) does not imply that versions form a universally linear history. Alternative evaluations may coexist, some corrections may be withdrawn, and a later release need not monotonically increase credibility for every application. The essential requirement is that a scientifically relevant revision can be inspected and linked to the evidence and methods motivating it.

Argo suggests practical governance requirements for global epistemic infrastructure. A user should be able to identify whether a record is preliminary or delayed-mode reviewed, recover the applicable data and metadata version when reproducibility requires it, and understand which corrections affect derived products. Providers need sustainable scientific labour and accountable processes for quality control. Contributors should receive recognition for observing and reviewing work even when the resulting data are openly distributed. A central catalogue, meanwhile, should avoid flattening multiple validity states into one undifferentiated notion of “current.”

There are limits to transferring this model wholesale. Argo works within carefully developed disciplinary conventions, mission-defined observing variables, and a relatively specialized technical community. Comparable standards for ecological knowledge, social records, culturally governed data, or AI-generated interpretations may be harder to define and may involve rights that make full publication inappropriate. The lesson is the value of institutionally supported, revisable knowledge, not a claim that the Argo workflow universally settles questions of authority or interpretation.

3.4 GOOS, OceanOPS, and Marine Coordination

The Global Ocean Observing System (GOOS) was established by UNESCO’s Intergovernmental Oceanographic Commission in 1991 and developed as a coordinated response to scientific and societal needs for ocean information. GOOS is presently led by IOC-UNESCO and co-sponsored by WMO, the United Nations Environment Programme, and the International Science Council [29, 31]. Its role differs from that of one autonomous observing array. It coordinates a plural set of instruments, scientific communities, operational requirements, and regional priorities. This makes GOOS useful for studying how a system can possess global aims without consisting of a single homogeneous data-producing organization.

A central coordinating mechanism is the designation of Essential Ocean Variables (EOVs), framed around observations needed to understand ocean processes and support public applications. GOOS distinguishes physical, biogeochemical, and biological or ecosystem-related variables and develops requirements connecting phenomena, observations, and uses [28]. This approach has two useful features. It connects monitoring priorities to research and operational questions, and it can provide continuity while particular instruments, platforms, and institutions change. At the same time, variable designation is an institutional and scientific judgment. What counts as essential can require revision as ecological knowledge develops, social needs change, or previously underrepresented communities identify neglected phenomena.

OceanOPS contributes a more operational form of coordination. The GOOS description identifies monitoring observing networks, exchanging platform metadata, supporting operations, and assessing system performance among its functions [30]. UNESCO-IOC also describes OceanOPS as a joint support capability for oceanography and marine meteorology whose work is sustained through contributions from participating programmes and states [39]. It helps participants locate operating platforms and gaps in coverage, while informing deployments and data exchange. These are not the same functions as making a scientific interpretation of every observed process or adjudicating disputes about commercial and community uses.

The contrast among GOOS, OceanOPS, and component networks offers a functional layering useful to later sections of this paper. GOOS supplies broader requirements and coordination; OceanOPS supports monitoring and operational metadata; component programmes such as Argo produce and review particular observations; forecasting and other scientific services use these observations in further infer-

ences. None of the components substitutes for all the others. If a float fails, the network's metadata may document a coverage gap. If an observation is corrected, the data stream should preserve the revision. If a particular ecosystem variable is poorly observed, a coordinating framework may identify the gap but cannot cure it merely by listing it.

An operational lesson concerns continuity. Data that are easily obtained during a short, well-funded campaign may be unsuitable for long-term trend estimation if maintenance and calibration end. GOOS institutional reporting in 2025 described both progress and persistent observational and sustainability gaps; the cited institutional page could not be directly inspected in the present verification pass, so the specific scope of its underlying measures remains to be checked against the report itself [38]. The existence of international coordination thus coexists with infrastructural precarity. Any proposed global knowledge commons would have to treat recurring maintenance, expert quality control, replacement equipment, energy, and long-term stewardship as parts of its economic and institutional constitution, not as marginal expenses after building a technical platform.

Marine observation also makes ecological and rights-related complexity especially visible. Ocean instruments can affect habitats or shipping and fishing activities; biologically sensitive information can create conservation risks; and the same observational record can be used for conservation, research, navigation, or commercial development. These uses do not necessarily share identical legitimacy conditions. A framework for global observing cooperation should therefore record the provenance and scientific meaning of marine data while specifying differentiated permissions and avenues for affected participants to raise concerns. A full assessment of marine-robotics ecology belongs to the companion ocean paper, whereas the present case establishes why environmental consequences belong in the governance of knowledge infrastructures.

3.5 Early Warning and the Translation of Observations into Action

International early-warning cooperation illustrates a limit of observation-centred institutional design. A technically credible hazard prediction can fail to protect people if exposure is not understood, a warning does not reach those at risk, or practical response is unavailable. The United Nations Early Warnings for All initiative organizes its programme around four linked pillars: disaster-risk knowledge; observation, monitoring, analysis, and forecasting; dissemination and communication; and preparedness and response [59]. The initiative describes distinct leads for disaster-risk knowledge (UNDRR), detection and monitoring (WMO), warning communication (ITU), and preparedness and response (IFRC). These announced responsibilities illustrate divided institutional functions, while their coordinated effectiveness in any particular setting requires separate evidence.

The four-pillar model underscores the separation of evidential quality, communicative accessibility, and legitimate action. An agency might correctly identify a storm trajectory while underestimating the vulnerability of an informal settlement. A warning might reach a telecommunications network but remain linguistically inaccessible or arrive without practical evacuation resources. A well-documented forecast could also trigger a disproportionate or contested intervention if authorities fail to communicate uncertainty or consult affected communities. Reliable knowledge is therefore a necessary input for some protective decisions but not, on its own, proof of an effective or just response.

The 2025 global multi-hazard early-warning report offers a useful empirical boundary to optimism. According to WMO and UNDRR, 119 countries, about 60% of the reporting global total, stated that they had a multi-hazard early-warning system; significant gaps persisted, particularly among small-island developing states [69]. This reports the existence of systems at country level, not the proportion of individuals effectively protected, the adequacy of local response, or independently verified uniform service quality. Consequently, the statement that global warning cooperation is a meaningful achievement should coexist with the finding that implementation remains uneven. The figures cannot be converted directly into a causal estimate of global protection without additional designs and evidence.

For a wider knowledge infrastructure, this case suggests a difference among at least four assessments: whether observations are appropriate; whether risk interpretations are warranted; whether warnings are

communicated to their intended recipients; and whether responses are feasible and legitimately governed. Interoperability between the first two does not ensure the last two. Such separation can be represented by a research-oriented trace from observations to decisions, with records of models, messages, intended audiences, responsible institutions, and subsequent corrections. A causal claim that one infrastructure reduced harms would require outcome evidence and attention to confounding factors; it cannot be inferred simply from the presence of these trace records.

There is also a constructive model for cooperation under incomplete trust. Actors can rely on specialized institutions that are accountable for limited tasks, use agreed terminology, and exchange records through verifiable procedures. Differences in jurisdiction and responsibility remain visible rather than being erased. Yet emergency coordination creates additional legitimacy questions about exceptional access, sensitive location data, responsibility for false alarms, and failures to warn. These questions require explicit rules, public scrutiny, and remedies; technical notification systems cannot settle them alone. Section 8 will examine these governance issues as part of an evolving relationship of reliance, contestation, and institutional repair.

3.6 The European Digital Twin Ocean and EDITO

The European Digital Twin Ocean brings the preceding infrastructures into conversation with a more explicitly computational project. The European Digital Twin Ocean is an ongoing initiative to connect marine data, models, simulations, and applications in support of research and decision-making. EDITO is its supporting public digital platform, integrating tools and resources drawing particularly on EMODnet and Copernicus Marine Service [20, 21]. This distinction between a broad knowledge-and-modelling ambition and its supporting platform matters. Neither the existence of an interface nor the operation of computational services establishes that the ocean is adequately represented at every location, ecological scale, or relevant decision context.

As described by the programme, the initial EDITO-Infra and EDITO Model Lab phase ran during 2022–2025, and the subsequent EDITO 2 project runs in 2025–2028. The stated ambition for an expanded, sustained European Digital Twin Ocean extends toward 2030 [20]. Current EDITO materials identify usable public platform components and services, including dataset and process catalogues and workflows involving near-data computation. September 2026 technical reporting describes progress in synchronizing EMODnet content and adapting resources for cloud-oriented access [19]. These are claims about specific reported components and development activities, not evidence that every proposed future service is fully operational or that the initiative already achieves comprehensive ecological representation.

The digital-twin approach adds an explicitly inferential layer. A measured ocean profile, a numerical circulation model, a satellite-derived surface product, an ecological distribution model, and a scenario projection are not interchangeable kinds of evidence. Simulations can integrate observations and expose alternative assumptions, but their outputs depend on model structure, boundary conditions, parameter choices, and validation. A decision interface might make these products easy to compare while obscuring their different uncertainties. The epistemic governance of a knowledge platform must therefore represent the difference between a measured input, an assimilated estimate, and a conditional scenario. It should also permit disagreement about a model’s purposes and the plausibility of proposed interventions.

The programme is especially useful as a test of what *interoperable computation* adds beyond interoperable datasets. Bringing algorithms toward shared data resources can reduce unnecessary copying and facilitate reproducible workflows. A federation might provide discoverable computational environments, versioned models, standard metadata and traceable results while retaining distinct data providers. However, computational access can introduce new inequalities: participants with expertise or resources to build and run models may exercise more influence over what counts as actionable knowledge. Public platform access alone does not guarantee equitable opportunities to create models, interpret outputs, or contest applications. Technical performance and institutional legitimacy must again be evaluated separately.

Table 2: Different institutional functions of international observation and knowledge programmes. Entries describe documented programme roles and analytic limitations, not comparative rankings.

Programme or mechanism	Transferable coordination practice	Main limit of analogy
WMO data policy / WIS2	Shared exchange classes; node publication; discovery and notifications	Domain-specific obligations do not authorize unrestricted use of all data.
Argo	Distributed observation with rapid and delayed expert quality control	Common variables and disciplinary review cannot be assumed across all domains.
GOOS / OceanOPS	Agreed observing priorities; platform metadata; monitoring of network coverage	Coordinating a network does not repair all coverage and financing gaps.
Early Warnings for All	Specialized institutions link risk knowledge, forecasts, communication and response	Warning production is not equivalent to protection or legitimate action.
European DTO / EDITO	Data-model-compute integration and collaborative simulation services	An evolving model platform is not a complete or authoritative representation of the ocean.

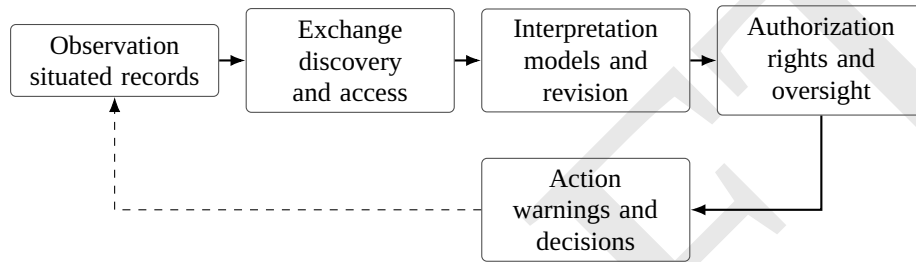


Figure 1: Analytical layers in a global observation-to-action process. Feedback indicates that practical outcomes can prompt renewed observation and critique. The ordering is explanatory rather than a single mandatory historical or operational sequence.

Ocean digital twins also provide an instructive boundary case for AI-assisted scientific inference. AI tools may suggest parameterizations, summarize observations, and support computationally intensive analysis. Their outputs can be valuable candidates for further inquiry without acquiring an independent right to override scientific scrutiny or community authority. If a digital twin informs marine planning, records of model versions, uncertainty, affected interests, and revision paths become relevant to decisions beyond the computational platform. Consequently, the EDITO case does not justify treating future globally integrated AI models as universal arbiters of knowledge. It does suggest concrete research tasks for accountable integration, verification, and collaborative modelling.

3.7 Transferable Mechanisms and Limits of Analogy

The cases above support an institutional comparison rather than a single generalization about global cooperation. WMO's exchange policy and WIS2 articulate rules and interfaces for distributed data movement. Argo illustrates jointly sustained observation and correction. GOOS and OceanOPS coordinate changing observing requirements, platforms, and operational metadata. Early-warning partnerships connect scientific assessment to communication and response, while EDITO explores the integration of models, data, computation, and interactive applications. These capacities overlap, but they are not identical. Table 2 records their principal transferable mechanisms and the limits of using them as precedents.

A minimal comparative design should therefore separate five layers: *observation*, *exchange*, *interpretation*, *authorization*, and *action*. An evaluation of one layer cannot stand in for the others. It should also distinguish programme declarations, evidence of deployed capabilities, independently observed performance, and judgments about governance or ethics. This is especially important when prospective benefits are used to justify public funding or the creation of indispensable digital services. Figure 1 provides a conceptual diagram of these distinct functions; the arrows denote possible informational dependence rather than an obligatory universal causal sequence.

The strongest transferable result is a possibility claim. Specialized organizations operating within

different political and scientific contexts can maintain shared observational arrangements when the required instruments, standards, institutions, and professional practices are sufficiently supported. The cases do not demonstrate that international rivalry has disappeared, that trust has been perfected, or that cooperation is equally beneficial for every participant. Indeed, international interdependence may co-exist with exclusion, unequal computational power, fragile funding, contested ownership, and uneven capacity to influence standards. These limitations are not incidental anomalies in an otherwise automatic convergence toward global scientific unity; they are part of the design problem.

A hypothetical global epistemic federation could borrow the modularity of WIS2, the revision histories of Argo, the monitoring coordination of OceanOPS, the action-sensitive accountability of early-warning systems, and the model-computation links being developed through EDITO. Yet combining these elements into a general infrastructure introduces new conflicts. Semantic translation across heterogeneous scientific and social domains can be noninvertible; access permissions can change; disagreement may be substantive rather than the result of missing metadata; and a highly influential AI mediation layer may centralize interpretive authority. These concerns require explicit interfaces among technical standards and governance procedures instead of the assumption that an inclusive platform architecture will automatically produce fair institutions.

Finally, international observation cooperation should be treated as a continuing collective practice whose conditions can be revised. It depends on the maintenance of sensors, archives, scientific labour, communication infrastructure, accepted standards, and arrangements for criticism. Some components may be sustainable through public funding; others may depend on research projects, voluntary contributions, or limited contracts. The observational system is itself environmentally and materially situated, and its effects on ecosystems and communities merit examination. The present section establishes comparative precedents and limits; Sections 4–10 will develop a candidate relational and federated architecture, then specify tests for interoperability, credible knowledge, governance, and ecological responsibilities. No single case above establishes that architecture in advance.

4 Observation as a Relational Infrastructure

This section specifies the observational layer on which an interoperable knowledge infrastructure would depend. Its purpose is to explain how scientific records originate in materially situated encounters, how measurement conditions constrain their interpretation, and how their credibility can be preserved or revised across institutions. The object is broader than a collection of sensors: an observation infrastructure includes the practices, calibration chains, representations, maintenance arrangements, uncertainty descriptions, and accountable records through which measurements become usable evidence. The discussion builds on the epistemic distinctions in Section 2 and the institutional examples in Section 3, while reserving the detailed federation architecture for Section 5, semantic translation for Section 6, and assimilation methods for Section 7.

The term *relational* has two specific roles. First, a datum is intelligible through relations among what is investigated, the observing arrangement, its reference standards, and the interests and procedures guiding interpretation. Second, observational records can be linked by derivation, comparison, calibration, correction, or contestation without presupposing that all participants inhabit a uniform epistemic or institutional position. Such links can be represented mathematically while leaving open questions about their actual scientific and social validity. In particular, a globally accessible record may be unusable for a given question, and an inaccessible or sensitive record may nonetheless be scientifically important.

4.1 Land, Ocean, Atmospheric, and Orbital Observation

Global observation is carried by a heterogeneous ecology of platforms. Land stations can maintain repeated measurements at instrumented sites, yet the placement and survival of stations depend on terrain, institutions, power, and maintenance. Oceanographic profiles and fixed moorings sample volumes and water columns inaccessible to many land-based methods; they are constrained by deployment logistics,

sensor drift, and communication intervals. Atmospheric observations include surface stations, radiosondes, aircraft, and remote measurements, whose effective supports differ in altitude, sampling frequency, and exposure to local conditions. Orbital sensors extend geographic reach, but typically produce remotely sensed radiances or other signals from which desired environmental variables must be retrieved under assumptions concerning radiative transfer, geometry, clouds, or surface properties.

These platforms need not measure identical objects even when their products share familiar names. An in-situ water-temperature estimate at a specified depth and a satellite-derived sea-surface-temperature product may differ in measurement support, retrieval method, reference convention, and uncertainty structure. Treating them as interchangeable because both contain the word “temperature” would substitute a lexical match for a defensible comparison. The Climate and Forecast (CF) metadata conventions provide technical mechanisms for recording coordinates, grid mappings, units, and standardized variable descriptions, but application-specific scientific comparability remains a further inquiry [7]. Observing requirements developed through the Global Climate Observing System (GCOS) similarly connect environmental variables to accuracy, stability, traceability, coverage, and scientific purpose [27].

Table 3: Distinct observational supports and illustrative comparability requirements. The examples describe potential constraints, not a comprehensive assessment of programme performance.

Platform family	Typical support	represented	Necessary interpretive qualifications
Terrestrial in-situ	Site, depth or height, installation and interval		Exposure, relocation, maintenance, local representativeness
Ocean in-situ	Profile, trajectory, mooring or discrete sample		Pressure/depth conversion, drift, delayed quality control
Atmospheric profile	Moving or ascending instrument and sampled column		Vertical support, changing position, instrument response
Orbital remote sensing	Footprint, scan geometry or gridded retrieval		Retrieval algorithm, viewing conditions, cloud and surface effects

An expanded network of observations can create complementary evidence without producing complete environmental knowledge. Ecological systems may include unrecorded populations, transient events, deep environments, and causal processes that the chosen observing variables are poorly suited to detect. Conversely, numerous measurements of the same easily accessible conditions can enlarge a database while contributing little to uncertainty in neglected regions. Programme coverage should therefore be assessed relative to specified inferential and public purposes, rather than by an unqualified count of sensors, files, or data volume. The distinction is consequential for later work on ocean exploration and autonomous observing robots: the object of inquiry is the ocean and its ecological relations, with infrastructure an instrument of investigation rather than a preselected industrial destination.

4.2 Measurement Events and Representational Boundaries

A measurement is an event in which some aspect of a system is engaged by an observing arrangement and recorded in an agreed representational form. Its record has multiple dimensions that a numerical value alone cannot retain. For the present investigation, we propose the following *descriptive schema*, whose components are typed placeholders rather than a universal measurement standard:

$$o_i = (\iota_i, y_i, \mathcal{S}_i, \kappa_i, \pi_i, v_i, U_i, P_i, \alpha_i). \quad (6)$$

Here ι_i is an identifier; y_i is the recorded signal or measurement representation; $\mathcal{S}_i$ is the claimed observational support (including location and any temporal reference when applicable); κ_i records instrument and calibration context; π_i denotes the observation protocol; v_i describes units, coordinate systems, and

semantic conventions; U_i is an uncertainty and quality description; P_i links to provenance; and α_i records access and authority conditions. The components of Equation (6) may refer to external governed records rather than requiring unrestricted publication of all associated metadata.

A simple conditional measurement relation is

$$y_i = \mathcal{M}_i(x|_{S_i}; \kappa_i, \pi_i) + b_i + \varepsilon_i, \quad (7)$$

where x denotes a proposed environmental state or phenomenon of interest, $\mathcal{M}_i$ is an observation operation, b_i represents systematic discrepancy relative to that operation, and ε_i represents unresolved variation or error. Equation (7) is a deliberately simplified analytic representation. It does not assert that every observation has an additive or linear error model, or that the putative state x can be specified without theoretical commitments. In practice, the measurand, calibration reference, filtering and processing steps, and even the relevant system boundary are themselves subject to scientific decisions [4, 67].

The word *event* must not be conflated with an enduring observer, person, sensor, AI system, or legal organization. In the related generative-relational programme, vertices can represent events, while durable agents, apparatuses, and institutions are identified through histories and relations connecting multiple events [35]. For this paper, a directed observational-dependency relation $o_i \prec_{\text{dep}} o_j$ means that the interpretation or production of o_j relies on some content or calibration of o_i ; it does not, by itself, establish a universal chronological ordering. Data acquisition may be indexed by UTC timestamps, local clocks, mission cycles, depth coordinates, or geological strata, but time synchronization is a contingent metrological and infrastructural achievement. When no defensible conversion exists, order may remain partial or unknown.

These boundaries matter when fusing data from independent programmes. A numerical match obtained after unit conversion can be invalid if the samples refer to different water masses, geographic supports, or instrument conventions. A record should therefore identify the support and transformations under which a comparison is claimed to hold, along with unresolved incompatibilities. The schema in Equation (6) provides a vocabulary for recording such claims; it does not make them true by definition.

4.3 Sensor Calibration, Quality, and Missingness

Observational credibility depends on traceability to relevant reference standards, suitably documented calibration, and repeated assessment of equipment behaviour. These are active practices: sensors age, fouling changes instrument response, reference standards are updated, and maintenance events can create discontinuities in a time series. WMO's measurement guidance treats instrument methods and traceability as part of operational meteorology; its October 2026 provisional edition remains subject to a later approval process [67]. An infrastructure spanning independently operated nodes needs to describe which calibration claims are accepted, for what quantities and ranges, and under which versions of the relevant procedures.

We distinguish a raw signal y_i , a calibrated representation $\tilde{y}_i$, and a quality description q_i . A simplified expression is

$$\tilde{y}_i = C_{\kappa_i}(y_i), \quad q_i = Q(\tilde{y}_i, \kappa_i, \pi_i, \mathcal{R}_i), \quad (8)$$

where C_{κ_i} is a stated calibration operation and $\mathcal{R}_i$ represents available reference or comparison evidence. Neither mapping in Equation (8) must be invertible, and a quality flag is not automatically a probability that the data are correct. Calibration uncertainty, instrumental bias, representativeness, and model-form discrepancy can have different inferential consequences; merging them into one universal scalar obscures their sources. When assumptions support a covariance approximation, correlated calibration errors across observations should be recorded rather than silently treated as independent noise.

Argo is instructive because its products distinguish rapidly available profiles from records subjected to detailed delayed-mode assessment and calibration. Programmatic guidance explicitly warns users of the importance of quality-control flags and adjusted variables; salinity drift and pressure biases can matter for analyses requiring high accuracy [55, 62]. The scientific meaning of a revised profile is

consequently not exhausted by replacing one number with another. Users also need to know the revised variables, quality reasoning, intended applications, and the identity of the data product version.

Missing observations require separate treatment. For a candidate measurement opportunity i , let $R_i = 1$ when a usable record enters the chosen analysis and $R_i = 0$ otherwise. Then

$$\Pr(R_i = 1 \mid x_i, z_i, \mathcal{I}_i) = g_i(x_i, z_i, \mathcal{I}_i), \quad (9)$$

where x_i denotes an unobserved relevant condition, z_i observable contextual factors, and $\mathcal{I}_i$ the infrastructure state. Equation (9) is a warning about selection, not an estimated probability model. Cloud obstruction, extreme weather, inaccessible waters, data-transfer failure, unstable institutional funding, or restricted access can systematically shape which records become available. An archive cannot establish that its unavailable observations are missing at random merely by recording a missing-data marker.

In a hypothetical marine observation network, strong storms may interrupt communication at precisely the times when extreme ocean processes are of interest. Filling these gaps with smooth model forecasts may be useful for some operational purposes, but the model-generated values should not be relabelled as direct storm-time observations. For a sensitive ecological survey, an absence from the public dataset might instead reflect legitimate restrictions on publishing a threatened species' location. Treating either case as a blank measurement with the same reason code would create both scientific and ethical problems. A practical system should distinguish non-measurement, failed transmission, unusable quality, lawful restriction, and deliberate noncollection when this can be done without disclosing protected information.

4.4 Spatiotemporal Sampling and Model Dependence

Observation infrastructures select where, when, and by what means evidence is acquired. A sampling design therefore establishes an *epistemic footprint*: some changes are resolved, others are averaged, and still others remain invisible. In the simplest discretized setting, let $x \in \mathbb{R}^n$ represent an environmental state under an explicitly selected model, and let $H : \mathbb{R}^n \rightarrow \mathbb{R}^m$ represent an observing operation. The resulting observation is written

$$y = Hx + \varepsilon. \quad (10)$$

The relation in Equation (10) is a local linear model suitable for illustrating identifiability; heterogeneous observations generally require nonlinear operations, coordinate transformations, and richer uncertainty models.

Proposition 4.1 (Non-identifiability under an observation nullspace). *Let $H : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be linear and suppose there exists a nonzero $v \in \ker H$. Even with noiseless observations $y = Hx$, the observation y does not uniquely determine x over $\mathbb{R}^n$.*

Proof. For every x , the distinct state $x' = x + v$ satisfies $Hx' = Hx + Hv = Hx$. Hence a single observation record is consistent with at least two modeled states. $\square$

Proposition 4.1 states a limited algebraic fact. Nonlinear noninjectivity yields an analogous ambiguity, but a model's physical constraints or additional prior information can sometimes rule out the proposed alternative states. Such constraints must be specified and evaluated; the observation alone does not warrant assuming them. The proposition also does not show that an actual environmental system possesses a finite-dimensional state of the stipulated kind. Its role is to prevent the existence of a shared dataset from being mistaken for the identification of every process contributing to it.

The problem is pronounced where sensing geometries differ. A point instrument samples conditions near a place and exposure; a swath sensor integrates or retrieves signals over a footprint; a float profile provides depth-indexed measurements along a trajectory; and a simulation product is conditioned on model discretization and data inputs. Resampling these products onto a common grid changes their representations but may leave gaps in scientific comparability. A shared coordinate frame can be necessary for interoperability while remaining insufficient for equivalence of measurement support. The

GCOS guidance accordingly emphasizes specifying spatial and temporal sampling, accuracy, stability, and reference observations in relation to monitoring objectives [27].

For purposes of global cooperation, observational sufficiency should be evaluated against declared questions. An observing system adequate for seasonal ocean heat-content estimation may not resolve local marine-habitat disturbances. More satellite passes do not necessarily resolve subsurface chemistry. An AI model trained on extensive historical grids may appear spatially comprehensive while remaining uncertain about poorly represented deep-sea processes. Consequently, the proposed infrastructure should describe its *knowledge limits* alongside its coverage maps, reporting questions or scales for which observations are informative and those for which further measurement or other forms of knowledge are required.

4.5 Scientific Interpretation and Contextual Uncertainty

The interpretation of an observation proceeds through scientific and practical choices. A value becomes relevant evidence only in relation to a question, a measurement model, accepted transformations, and warranted uncertainty claims. Observation sharing can improve collective interpretation by allowing independent reanalysis and comparison, but it can also transport unexamined assumptions across contexts where they are poorly suited. This is a reason for preserving interpretive provenance, rather than an argument for insulating observations from criticism.

For a proposed scientific model M , the likelihood notation

$$p(y \mid x, M, \kappa, \pi) \quad (11)$$

expresses assumptions about how Equation (7) may produce observations. The expression in Equation (11) does not supply a posterior distribution, uncertainty calibration, or uniquely preferred interpretation by itself. Any inference also depends on assumptions about the modeled state, priors or constraints, discrepancy, and the relevant observational selection mechanism. A model ensemble may expose certain structural disagreements, but cannot guarantee that every relevant interpretation has been represented.

Table 4: Illustrative uncertainty layers whose evidence and responsibilities may differ.

Layer	Question for interpretation	Possible documentation
Instrument and calibration	What signal was measured and against which references?	Calibration chain, drift tests, validity ranges
Sampling and missingness	Which regions or conditions enter the record?	Coverage, exclusion reasons, selection assumptions
Retrieval and modeling	How was a physical quantity inferred from a signal?	Operator version, validation and discrepancy tests
Semantic and contextual	Do matching fields represent comparable quantities?	Support, units, definitions, purpose and known losses
Revision and lineage	Which records and results depend on a changed input?	Version identifiers, derivations, affected products
Rights and authority	Who can inspect, amend, restrict or contest an interpretation?	Custodial roles, access rules, review and remedy routes

Table 4 is not a proposal to convert governance disputes into statistical variances. Calibration uncertainty can sometimes be given a probabilistic account; the denial of a community’s authority to describe its own observations may instead be a matter of participation, interpretation, or rights. A single confidence interval cannot resolve both. Nor does an open file automatically satisfy an affected community’s concern about how its observations are interpreted or used. These distinctions extend the attention to situated knowledge, epistemic criticism, and rights developed in Section 2 [32, 47, 6].

A useful boundary test is an apparent disagreement between two ocean-temperature datasets. One record may represent a calibrated in-situ measurement at depth, another a satellite retrieval over a surface footprint. The values need not agree even when both systems operate correctly. Resolution could involve

comparing measurement supports, stratification, sampling times, retrieval assumptions and uncertainties; selecting a preferred dataset solely by institutional prestige would evade the scientific question. At the same time, when affected communities identify harms from data publication, purely scientific reconciliation does not exhaust the obligations of custodians and users. The infrastructure should make room for both epistemic criticism and legitimate claims about use.

4.6 Corrections, Revisions, and Provenance Histories

Science frequently advances through correction: instruments are recalibrated, processing algorithms are improved, suspect profiles are reclassified, and revised evidence changes earlier results. A distributed infrastructure therefore needs more than a current-value database. It needs records of how a scientific claim was generated, what it used, which later version superseded it, and which interpretations remain affected. The W3C PROV ontology offers an established vocabulary of entities, activities, agents, derivations, and revisions for provenance representation [71]; here we use it as one conceptual and technical reference, not as a complete semantics of scientific justification.

Let $\mathcal{V}$ denote a set of versioned observational products, and let $\prec_{\text{rev}}$ denote an asserted derivation or revision dependence. One may store a directed revision graph

$$P = (\mathcal{V}, E_{\text{rev}}), \quad E_{\text{rev}} = \{(v_i, v_j) : v_i \prec_{\text{rev}} v_j\}. \quad (12)$$

An append-only implementation can require Equation (12) to be acyclic by assigning a new version identifier to each revision and disallowing dependencies on versions that are not yet established. This is a record-management constraint, not an assertion that every scientific discovery occurs in a simple temporal sequence. Independently produced products may have no ordering relation; competing interpretations can branch, remain unresolved, or be incorporated into a later synthesis. The archive should preserve such differences rather than silently replace older claims with a supposedly universal final interpretation.

A compact illustrative history may proceed as follows. Product v_0 contains an automatically quality-controlled ocean profile. A laboratory or expert review identifies a possible salinity drift, producing a documented adjusted product v_1 with its calibration evidence. One downstream model run m_0 used v_0 ; another, m_1 , uses v_1 . A later user can distinguish the two inferences and decide whether the difference is consequential for the declared application. The example corresponds to an established style of staged Argo quality control, but the version identifiers and downstream models are hypothetical [55, 62].

Provenance is necessary for many forms of accountability, but it cannot independently certify that a sensor was properly calibrated, a model was valid, or an institution exercised legitimate authority. A perfectly traceable chain of mistaken assumptions remains mistaken; a chain containing sensitive information may also require controlled disclosure. Accordingly, the infrastructure should support both the *availability of appropriate provenance* and the *capacity to challenge provenance claims*. This includes access to correction mechanisms, explicit reasons for withdrawal or restricted release where disclosure is permissible, and preservation of audit evidence under suitable institutional controls. FAIR-oriented discoverability and reuse can be combined with differentiated authority and the CARE principles where relevant, rather than being interpreted as a mandate to disclose every record [61, 6].

The six dimensions developed in this section lead to a bounded conclusion. Observational interoperability requires jointly interpretable records with stated supports, calibration context, missingness, uncertainty, revision and authorization histories. It does not require that all participating systems agree on a single model or surrender control of their records to a central institution. The following section will investigate how federated nodes, registries, access services, and accountable computational interfaces might preserve these conditions at scale. The independent substantive questions of semantic translation, assimilation, and transnational governance remain for their dedicated sections and companion research papers.

5 Federated Global Observation Architecture

The previous section established why scientific observation requires contextual records, traceable transformations, and the possibility of correction. This section specifies an exploratory infrastructure architecture in which these requirements might be preserved across independently administered systems. Its aim is neither to replace existing meteorological and oceanographic arrangements nor to prescribe one uniform platform. Instead, the proposed federation specifies the minimum interfaces through which institutions with different instruments, resources, mandates, and rights can publish discoverable descriptions, negotiate access, conduct limited analyses, and review the resulting knowledge claims. The architecture is a research proposal: no claim is made that its composite design has been deployed, evaluated, or authorized by the programmes discussed in Section 3.

The distinction between a distributed set of machines and a *federation* is substantive. Machines can exchange packets without their operators recognizing any shared commitments. Conversely, institutions may coordinate standards and responsibilities while maintaining physically separate repositories, computing facilities, and scientific practices. A federation requires specified interfaces and a continuing practice of resolving incompatible records, operational failures, and claims of authority. We therefore represent an infrastructure as a configuration of technical functions *and* governed relations, rather than as a single repository with geographically dispersed mirrors. WIS2 provides a bounded operational comparison through its nodes and global discovery, notification, cache, and monitoring functions [64, 65]. The evolving European Open Science Cloud (EOSC) federation provides a complementary instance of institutional and service federation, but its participation rules, geographic remit, and maturity do not warrant extrapolation to all global knowledge domains [14].

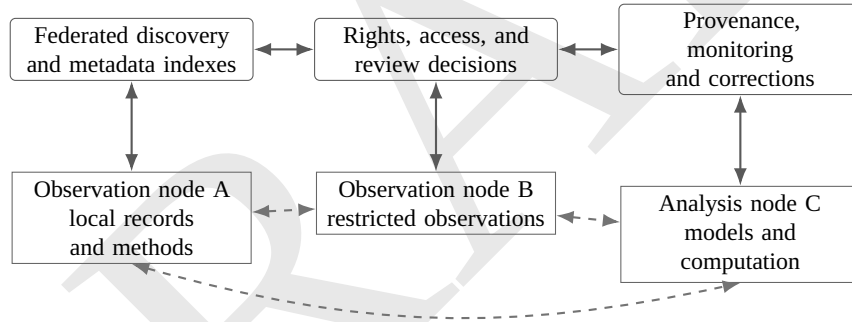


Figure 2: Conceptual federation, not a deployed system. Solid connections represent governed discovery, authorization, or notification interfaces; dashed connections denote conditional data or computation exchanges. Nodes can be institutions, services, or jointly administered facilities, and roles may overlap without implying shared ownership.

Figure 2 separates publication and discovery, the decision to allow a specified use, and the performance and correction of that use. None of these functions alone guarantees scientific validity or institutional legitimacy. An authorization decision, for instance, may determine whether a researcher can retrieve a restricted dataset; it cannot determine whether the retrieved variables measure a comparable phenomenon. Conversely, two measurements can be comparable without their custodians having permission to exchange them. The following subsections develop the resulting interfaces and their failure conditions.

5.1 Nodes, Participants, Roles, and Responsibilities

A node is a governed point of participation that may expose one or more services: an observing facility, an archive, a metadata catalogue, a model repository, a computational service, or a coordinating institution. The entity responsible for a node should be distinguished from the physical machines implementing it. A university may maintain instruments on behalf of an international programme; a government agency

may publish records through a contracted cloud operator; a research consortium may jointly govern an archive while individual investigators retain responsibility for calibration. These different configurations give rise to different obligations for accuracy, security, service continuity, scientific interpretation, and remediation.

Let $\mathcal{N}$ denote a set of registered service nodes, $\mathcal{I}$ a set of potentially accountable institutions, and $\mathcal{F}$ a set of functions including observation, custody, discovery, computation, authorization, and correction. The proposed role assignment is a relation

$$\mathcal{R}_{\text{role}} \subseteq \mathcal{I} \times \mathcal{N} \times \mathcal{F} \times \mathcal{H}, \quad (13)$$

where $\mathcal{H}$ denotes represented institutional histories or validity contexts rather than a universal clock. A tuple in Equation (13) documents that an institution undertakes a function with respect to a node under stated conditions. It is not an ontological claim that institutions are primitive vertices in a relational spacetime; their identities and responsibilities can themselves be represented as historically constituted relations and recognized practices. Role records require versioning as administrators, mandates, contractors, and permissions change.

At minimum, the architecture should record responsible custodians, scientific stewards, operational contacts, applicable rules, and mechanisms for disputed decisions. A scientific steward may assess calibration and model interpretation but have no authority to grant disclosure; a data controller or corresponding lawful decision-maker may authorize specified access without certifying the scientific claim. Contractual technical operation and public accountability need similarly separate representation. Some nodes may be small and resource-constrained; membership criteria that demand expensive continuous service could reproduce epistemic exclusion. A graded participation model, including intermittently connected and community-operated observatories, therefore deserves evaluation alongside high-availability institutional nodes.

One failure mode arises when institutional relationships are mistaken for infrastructure identities. If two nodes share an operator or funding source, their failures and decision-making may be correlated even when they appear as distinct endpoints. A meaningful federation inventory should therefore describe jurisdiction, governance independence, funding and maintenance dependencies where disclosure is appropriate, and critical upstream services. Network topology alone cannot establish resilience or institutional pluralism.

5.2 Discovery Services and Federated Registries

Global observation is not served by a central registry containing every record and measurement. A discovery system can index descriptions of datasets, services, models, versions, and access conditions while leaving the underlying holdings at institutionally governed sites. OGC API-Records Part 1 establishes a web-oriented means of exposing searchable metadata collections, and WIS2 provides an operational example of discovery metadata disseminated across participating nodes [50, 65]. These examples show particular interfaces; neither establishes that a record located through discovery is scientifically adequate, legally reusable, or currently obtainable.

Consider a resource d with a description $\mu(d)$, a declared endpoint or resolver $\ell(d)$, a version identity $v(d)$, and an access-policy reference $a(d)$. A discovery response may expose

$$\delta(d) = (\text{id}(d), \mu(d), \ell(d), v(d), a(d), \sigma(d)), \quad (14)$$

where $\sigma(d)$ records a publication, integrity, or freshness status. Equation (14) is an analytical record proposal, not a new OGC or WMO standard. Discoverability may refer to the existence of an otherwise protected resource without exposing its contents or sensitive location. In ecological monitoring, even precise occurrence coordinates or the identities of vulnerable communities may be sensitive; discovery metadata can therefore require redaction or coarsening.

A federation should distinguish the authority that asserts a record, the service that indexes it, and the observer who later uses it. Persistent identifiers, as supported through data citation practices and

the DataCite metadata schema, facilitate attribution and retrieval, but a persistent identifier is neither an assurance of permanence nor a substitute for scientific content review [12]. Registries must support invalidation, duplicate reconciliation, changes of custodianship, version redirection, and records that remain accessible only as withdrawal notices. Delayed synchronisation can be acceptable for stable metadata yet dangerous for emergency alerts, revoked permissions, or corrected calibration data; those classes need different update requirements.

A useful discovery contract therefore specifies query semantics, provenance of catalogue assertions, maximum intended staleness, identifier resolution behaviour, error responses, and a process for challenging mistaken records. Federation success should be measured not merely by the number of indexed datasets but by the ability of diverse users to locate relevant, interpretable, appropriately governed resources. Later empirical work should test false negatives, inconsistent indexing, incomplete provenance, multilingual semantic mismatches, and unequal discoverability between well-funded and small contributors.

5.3 Metadata, Identity, and Provenance

The observation tuple introduced in Equation (6) distinguishes measured content from the conditions of its production. Federated metadata must preserve that distinction when a dataset is copied, reformatted, quality-controlled, or interpreted at another node. In particular, a name or file checksum identifies a digital object relative to specified conventions; it does not identify the complete physical event, establish calibration correctness, or reveal every intermediary transformation. The W3C PROV-O recommendation provides a reusable vocabulary for entities, activities, agents, and relationships among them [71]. We use it as a possible interoperability resource, not as evidence that encoding provenance guarantees scientific truth.

For a derived record d' , an auditable account should permit inquiry into the source records, transformation operation, software and calibration versions, institutionally accountable actors, and output uncertainty. We propose a versioned directed graph $\mathcal{P} = (V_P, E_P)$ whose vertices include *record versions and activities* and whose edges assert typed derivation, usage, generation, or correction relations. The graph is an epistemic representation and may omit facts that were never observed or are legitimately restricted. PROV-O agents are modelling categories for attributed responsibility; their use does not imply that an AI service possesses legal personality, moral standing, or an independent right to authorize access.

Provenance records may themselves create safety, privacy, and governance risks. For example, revealing a detailed instrument-deployment path may reveal the location of endangered species or sensitive infrastructure. Thus an architecture may expose a public trace of derivation while preserving detailed audit evidence within an authorized trust domain. A cryptographic digest can help detect unexpected changes in an accessible byte sequence; it cannot independently certify that the original measurement or research methodology was sound. Likewise, a digitally signed record attributes a claim to a key under an identity arrangement, but the validity of its institutional mandate and the quality of its measurements require other forms of assessment.

Identity must be specified across at least four domains: resources and versions; responsible people or organizations; technical processes and services; and permitted uses. Conflating these domains permits a service credential to be mistaken for scientific authorship, or a citation identifier to be mistaken for a disclosure license. Revision histories should distinguish correction of a faulty record, changed scientific interpretation, and changed access authority. The resulting account need not force all knowledge histories into a universal total ordering: dependencies among observation, analysis, and revision can often be represented as a partial order, together with local timestamps and explicit synchronization evidence when available.

5.4 Data Access, Authorization, and Rights Management

A federation cannot infer a universal right to use data from technical reachability. Public environmental observations, confidential research records, community-controlled knowledge, security-sensitive locations, and licensed model outputs present different questions. Authorization should be evaluated for a particular request, declared purpose, applicable rules, claimant, and institutional context. For example, GA4GH’s Data Use Ontology encodes standardized permitted-use terms in the genomics domain; its existence demonstrates one method of making use restrictions machine-interpretable, not a universal solution to environmental or transnational rights [25].

Let p be an authenticated requester, d a governed resource, u a stated use, c an operative institutional context, and h a represented authorization history. A proposed access decision is

$$\text{Allow}(p, d, u, c, h) \in \{\text{allow}, \text{deny}, \text{review}\}. \quad (15)$$

The third outcome in Equation (15) represents unresolved cases requiring accountable review, rather than silently treating legal or ethical uncertainty as permission. The function is *a model of a decision procedure*, not an assertion that any automated classifier can substitute for valid consent, law, rights-holder authority, or democratic governance. Permission to discover that a resource exists, to inspect its metadata, to compute a protected aggregate, and to download the underlying observations may be separately granted or prohibited.

Authorization also has lifecycle conditions. Permissions may be limited by purpose, duration, geography, contractual constraints, community authority, or emergency exceptions; revocations must reach caches and delegated processors under stated guarantees. Complete withdrawal may conflict with lawful retention, independent reproducibility, or downstream publications, and it may be technically impossible to retract knowledge already inferred from legitimately accessed data. Governance should distinguish these cases instead of promising perfect deletion. Access logging can improve accountability, but logging itself may expose sensitive activity and therefore requires its own access controls and retention policy.

Federation members should be able to dispute incorrect classifications, demand explanations appropriate to the decision, and obtain review by a responsible authority. Equity also matters: institutions with little legal or technical capacity may face disproportionate burdens in negotiating permissions. Standardizing machine-readable conditions is useful only when coupled with comprehensible processes for contributors and affected communities. The present paper specifies these functions as design requirements while deferring substantive jurisprudence and cross-border adjudication to the companion governance investigation.

5.5 Storage, Replication, and Resilience

A distributed system may fail at many levels: the originating sensor can cease to function, the local archive can lose power, a registry can become unavailable, an international link can fail, or a policy decision can make a cached copy unusable. Resilience therefore concerns continuity of observing capacity, preservation of scientific records, reliability of metadata and permissions, and the ability to reconstruct dependencies following an outage. Replication improves some forms of availability but may simultaneously multiply ecological costs, unauthorized disclosures, version inconsistencies, and dependencies on a small set of cloud providers.

Under a deliberately restrictive toy model of independent replica availability events with probabilities $p_1, \dots, p_k$, the probability that at least one replica is reachable is

$$\Pr(\text{at least one reachable}) = 1 - \prod_{i=1}^k (1 - p_i). \quad (16)$$

Equation (16) is not a prediction for real scientific infrastructure. Its independence assumption is frequently false when replicas share power grids, legal restrictions, administrative credentials, upstream

networks, or environmental hazards. High nominal availability can also conceal a lack of fresh measurements or trustworthy calibration, so *reachable bytes* should not be equated with *usable current evidence*.

Table 5: Distinct resilience objectives in federated scientific infrastructure. Each measure requires an explicitly defined observation window and operational setting.

Function	Failure to evaluate	Illustrative evidence or mitigation
Observation continuity	Instruments remain offline while servers are available	Maintenance logs, redundancy of sampling, coverage gaps
Discovery continuity	Metadata becomes stale, unreachable, or misleading	Replica lag, catalogue tests, withdrawal propagation
Record integrity	Files change without valid version or correction history	Checksums, immutable audit records, provenance review
Authorized use	Cached copies bypass changed access decisions	Delegation expiry, cache invalidation, incident response
Independent recovery	Common provider failure disables multiple nodes	Dependency maps, restoration exercises, portable formats
Ecological burden	More replicas increase material and energy impacts	Lifecycle inventories, marginal energy and material costs

Table 5 indicates why the appropriate replication strategy depends on scientific criticality, geographical and institutional independence, data sensitivity, cost, and legitimate retention. A core public emergency-warning feed may justify rapid redundant distribution; some biodiversity records may require restricted geographically separated custody; large derived datasets may be reproducible more efficiently from preserved methods and inputs. All strategies involve maintenance obligations and should be evaluated against both resilience gains and additional environmental burdens. These are conditional design possibilities, not claims about universally optimal engineering choices.

5.6 Computational Facilities and Cross-Border Analysis

Cross-border knowledge production may require bringing algorithms to governed data rather than transferring all records to a common archive. An observation custodian might permit execution of an approved analysis within its environment, release appropriately checked summary results, or contribute model updates to a distributed computation. Other investigations may require raw records for calibration review, replication, or the diagnosis of systematic disagreement. Thus *data locality* is one useful architectural option, not an ethical guarantee or a universal replacement for transparent scientific access.

We describe a requested computational action by a typed manifest

$$j = (I, M, C, U, Q, O), \quad (17)$$

where *I* denotes declared input record versions, *M* model and software identities, *C* computing conditions, *U* authorized purpose, *Q* output-review constraints, and *O* the proposed output representation. A valid execution requires both technical compatibility and institutionally authorized input and output operations. Equation (17) supplies a reporting structure for scientific reproducibility; it does not define an executable universal API. Different nodes may implement distinct secure environments, batch processing, common scientific workflows, or explicitly negotiated cross-site analysis agreements.

Even when raw data never leaves a custodian, query responses, aggregates, model parameters, timing information, or repeated interactive requests may disclose sensitive information. Assertions that *computation follows the data* must therefore be tested against explicit threat models, output governance, and methodological needs. Furthermore, unequal computing power can shift interpretive authority toward those who can run large simulations, maintain models, and publish consequential outputs. A federation's equitable participation requires attention to compute allocations, energy access, methodological review, scientific training, and support for independent replication.

A scientifically meaningful execution record should preserve input versions and calibration status, software and model specifications, environment information relevant to reproducibility, output transformations, stated uncertainties, and reviewers' correction decisions. Some technical details may be restricted for legitimate safety or privacy reasons; a defensible substitute should explain the limitation and enable an appropriate controlled audit. Section 7 will investigate assimilation and model discrepancy in detail. For present purposes, the architecture provides a way to stage, authorize, and critique computation across institutional boundaries while admitting that independent analysis and computational authority are distributed unequally.

5.7 Model Registries, Auditability, and Correction

Observation records do not become knowledge simply through aggregation. Scientific claims depend on models of the measurement process, assumptions about the observed systems, methods of error correction, and interpretation within a research question. Model registries should therefore be connected to observation provenance without conflating the two. A registry entry ought to specify model purpose, version and provenance, inputs and observational supports, calibration or training regime as relevant, known limitations, evaluation setting, and the institution responsible for reviewing reported defects.

A useful distinction is between *artifact identity*, *epistemic claim*, and *institutional endorsement*. A model file can be intact while its assumptions are inappropriate for a new ocean basin or climate regime. A forecast can be produced by an authorized institution yet remain uncertain or later require correction. A model update can improve performance for one application while reducing it for another. Consequently, the federation should permit parallel models, competing uncertainty statements, and explicit disagreement about interpretation. It should not silently select one universally privileged world model merely because one system can compute more outputs.

Proposition 5.1 (Registry agreement does not establish epistemic agreement). *Suppose two nodes publish metadata records that agree on a resource identifier, version, nominal variable name, and access endpoint. Without further assumptions on observational support, calibration, units, semantics, and inference procedures, this agreement does not imply that their derived scientific claims are equivalent.*

Proof. Take a shared, correctly identified satellite radiance record. One node applies a retrieval model using one set of atmospheric and surface assumptions, while another applies a different defensible model and obtains a distinct temperature inference. Both nodes can cite the same record identifier, record version, and access endpoint. Those shared identifiers do not establish equivalence of the derived scientific claims. Additional assumptions about their inference procedures and validation domains would be necessary. $\square$

Proposition 5.1 is a counterexample to an overly strong interpretation of federation interoperability. It does not show that the two observations are necessarily incompatible, only that their compatibility must be evaluated. Corrections must likewise propagate across levels: a calibration revision can change dataset versions, invalidate a derived product, or trigger an amended public interpretation. Notification should carry both machine-readable dependency information and a meaningful route through which affected researchers can contest the proposed correction. A signed provenance graph can document what happened; the decision whether a revision is scientifically justified remains a substantive inquiry.

A transparent model registry can improve accountability but cannot guarantee the absence of hidden training data, biased representations, opaque institutional incentives, or undisclosed environmental costs. For AI systems in particular, a model's ability to assimilate heterogeneous streams should not be mistaken for unlimited knowledge of the underlying world's generative structures. Section 6 will examine translation and epistemic non-equivalence; Section 7 will examine model pluralism and propagation of uncertainty.

5.8 Cross-Layer Mediation and Review Contracts

The preceding services cannot be joined merely by passing resource identifiers among catalogues, analysis nodes, translation agents, and institutional decision portals. A federation requires a *cross-layer contract* describing what a requested operation consumes, what it may produce, which dependencies remain inspectable, which reviewers can assess its claims, and which authority, if any, can permit the proposed use. The contract is a research application profile. It is not an enforceable universal API, nor does its presence establish shared meanings or governance legitimacy. Its purpose is to prevent an inference valid in one service layer from silently becoming a purported permission or scientific conclusion in another.

For a bounded operation z and documented configuration h , define an attributed review profile

$$\mathfrak{R}_h(z) = (\rho_{\text{tech}}, \rho_{\text{epistemic}}, \rho_{\text{authority}}, \rho_{\text{challenge}}; \mathcal{E}, \mathcal{P}). \quad (18)$$

Each ρ is an independently attributable assessment concerning, respectively, technical interchange, task-specific scientific or interpretive adequacy, competent permission for the requested operation, and availability and outcome of relevant contestation. The evidence register $\mathcal{E}$ retains scopes, assumptions, counterexamples, decision sources and assessor identities; $\mathcal{P}$ retains versioned provenance and known dependency references. A coordinate may be recorded as supported, refuted, unresolved, or unassessed *according to a declared assessment procedure*. These terms are not interchangeable with the preservation-test statuses in Equation (48), institutional acceptance in Equation (40), or the policy responses illustrated in Equation (76). In particular, a positive scientific assessment supplies no inference to permission, and an authenticated policy assertion supplies no proof of legal competence. For each layer ℓ , the assessment should carry a named assessor a_ℓ , an operation-specific claim c_ℓ , an evidence boundary D_ℓ , and a disposition from $\Sigma = \{\text{supported, refuted, unresolved, unassessed}\}$. Different assessors may issue incompatible dispositions on the same claim; Equation (18) denotes one attributed profile or collection of profiles, not a globally authoritative aggregation algorithm. The four status words describe evidentiary dispositions and never silently replace the differently typed policy outcomes permit, deny, and undetermined.

Table 6: Cross-layer hand-offs and explicit boundaries in a proposed federation. Each output remains conditional on its own declared review and institutional context.

Hand-off	Evidence carried forward	Claim that cannot be inferred solely from receipt
Observation to discovery	Source/version, calibration support, disclosure scope	Indexed observations constitute validated scientific claims
Discovery to research	Input support, execution envelope, applicable use conditions	Accessible bytes ensure reproducibility or authorized execution
Research to mediation	Interpretive question, models, uncertainty and counterexamples	A fluent mapping preserves all relevant distinctions
Mediation to governance	Bounded candidate, lost meanings, participants' reservations	Translation approval confers a right to use or publish
Decision to circulation	Competent review source, allowed acts and dependencies	One permission covers new purposes, derivatives or jurisdictions
Correction to dependants	Revision event, known dependency reach and recipients	Notifications erase all copies or settle disputed validity

The dependency register needs typed links that distinguish *used*, *derived from*, *interpreted under*, *reviewed by*, *authorized for*, and *challenged by*. These are statements attributed to identified sources, not interchangeable causal relations and not the vertices or edges of a presumed ontological two-complex. A

correction query can follow declared derivation links to identify known affected products; this gives neither a completeness guarantee for unregistered copies nor an automatic right to cancel a legally preserved record. If a mediation path passes through two independently reviewed mappings, their scopes, supports and rights must be rechecked for the composed purpose. A path of individually well-typed mappings is not by itself a validated composite: the first output must lie within the second mapping's tested domain, the two review contexts must justify using comparable target quantities, and a direct mapping, when supplied, must be compared against the path result. Appendix C provides a conditional composition bound; Appendix A gives a counterexample to untested path consistency. A failed or unresolved authority assessment should block any *claimed authorization by the federation*; whether underlying actors may proceed is a question for applicable competent processes.

The architecture is thus operationally plural. One node can keep protected data local and expose a bounded, reviewable result; another can publish a conflicting model or withhold a proposed correspondence. Participants may disagree about both a mapping and the forum competent to decide its use. The federation should preserve such attributed differences without resolving them by a default central assessor. Section 6 examines the limits of the interpretive part of this contract; Section 8 treats its institutional conditions, and Section 10 extends it to common knowledge creation and value circulation.

5.9 Institutional Interfaces, Exit, and Accountability

Federation is sustained through institutions as well as protocols. Participants must have a way to negotiate standards, revise operational requirements, contest discovery or authorization decisions, investigate incidents, and exit arrangements that cease to serve their legitimate interests. A node's technical capability to disconnect cannot settle how other participants should treat previously exchanged records, outstanding obligations, emergency warning dependencies, or licensed downstream derivatives. Entry and exit require public-facing procedures whose consequences can be reviewed, including obligations that survive withdrawal and the circumstances under which access may be restricted.

Table 7: Proposed architectural responsibilities and minimum accountability questions. These are research-design commitments, not descriptions of rules already accepted by international organizations.

Interface	Responsible type	Question requiring accountable resolution
Node enrollment	Membership and operational eligibility	Who may join, on what terms, and with what support?
Discovery and identity	Publication, attribution, and correction	Who may challenge misleading or missing metadata?
Access and permitted use	Authorization and review	Whose rights or mandates govern a proposed use?
Cross-border computation	Execution and output release	Who bears responsibility for harmful or invalid outputs?
Incident and withdrawal	Containment, continuity, and exit	Which copies, dependencies, and duties survive a departure?
Scientific interpretation	Review, disagreement, and correction	How can alternative interpretations and errors be heard?

Table 7 exposes a limitation of a purely technical federation proposal: every interface presupposes an allocation of authority that may be disputed. An open technical standard does not dictate whether admission is equitable; open-source implementation does not decide ownership; distributed storage does not confer political legitimacy. A workable research design should therefore evaluate responsibility as a function of specific operations, recorded decisions, applicable rules, and the position of affected parties. The companion trust and governance paper will investigate how such rules might arise, change, fail, and be repaired rather than treating institutional trust as a static parameter.

We can express the distinction between formal service ability and legitimate action without collapsing

them into a single optimization score. For an operation o and governed context c , write

$$\text{Executable}(o, c) \not\Rightarrow \text{Authorized}(o, c) \not\Rightarrow \text{ScientificallySound}(o, c). \quad (19)$$

The non-implications in Equation (19) require no universal scoring rule: an operation may execute without permission, and an authorized operation may implement a mistaken model. The converse implications also fail in general. Institutional arrangements should make these questions independently auditable. Emerging implementations, such as the EOSC federation of scientific nodes, provide relevant experience in negotiated interoperability and organizational coordination but are not sufficient evidence of global participation, equal influence, or universal ethical adequacy [14, 18].

The architecture in Figure 2 is consequently a collection of interfaces open to empirical testing, not a vision of frictionless unification. Its performance must be examined under network partitions, mismatched vocabularies, permission changes, missing provenance, correlated provider failures, disputed data rights, and divergent institutional interpretations. The next section will turn to epistemic interoperability and generative representation, asking what it would mean for heterogeneous observational systems and models to translate one another’s claims, and which forms of loss, noninvertibility, and disagreement remain even when the underlying technical services function as designed.

6 Epistemic Interoperability and Generative Representation

The preceding section specified a federation of observation custodians, registries, computing facilities, and institutional interfaces. Such an architecture can identify data resources, transport authorized records, and document the transformations of scientific artifacts. The present section examines a further problem: *what would make observations, claims, and models produced by heterogeneous systems mutually interpretable?* An observation system has methods for deciding what counts as an object of measurement, how a signal becomes a datum, which inferential operations are admissible, and under what conditions a claim may be revised. Exchanging the resulting data does not necessarily exchange this generative context. Conversely, differences in method do not always prevent useful, explicitly bounded comparison.

We distinguish syntactic exchange, semantic alignment, measurement comparability, inferential translation, and institutionally authorized interpretation. The inquiry is partly informed by an earlier exploratory investigation of *knowledge pressing*, which asks whether one system can infer another’s generative representation rather than merely translate its outputs [35]. That earlier investigation is an intellectual precursor, not evidence that such inference can be performed completely or reliably. Here the narrower research object is the design of interoperable global observation infrastructures, including a way to represent what translation preserves, loses, or leaves disputed. The account remains provisional and does not claim to implement a general translation algorithm. Formally, a task-scoped map between record spaces (Appendix C) is a narrower object than a proposed correspondence between generalized languages, which may concern operations, interpretive histories and practical authority. Existence of the first supplies neither existence nor uniqueness of the second. The term *translation* is accordingly typed by its source, target, purpose, relation kind, and applicable review context; an exchange success at one type is not automatically transportable to another.

6.1 Syntactic and Semantic Interoperability

Syntactic interoperability makes independently produced records parseable according to agreed encodings, schemas, data types, interfaces, or validation rules. One service may deliver a time-indexed array; another may publish geospatial features; a third may expose observations through an API. Successful parsing guarantees that a recipient can read the stipulated structures. It does not guarantee that the array elements, feature properties, or model outputs refer to comparable phenomena. Structural conformance, metadata completeness, and scientific relevance should therefore be evaluated independently.

Semantic interoperability introduces controlled terms and relations for specifying observable properties, features of interest, sampling procedures, units, and the meaning of transformations. The W3C/OGC

SOSA–SSN ontology, for example, offers a vocabulary for observations, sensors, observed properties, and features of interest; the OGC SensorThings API offers an interoperable model and interface for retrieving observations and associated descriptions [75, 49]. These standards address important description and exchange requirements. Neither constitutes a universal theory of the phenomena under observation, nor does use of a shared label establish measurement equivalence.

An illustrative semantic record can be written as a typed tuple

$$\sigma(o) = (F, P, U, S, M, K, V), \quad (20)$$

where F denotes the feature or target, P the property to be represented, U the physical unit and reference frame where applicable, S the observational support or aggregation domain, M the method and sampling context, K the calibration and quality information, and V the relevant record version. The tuple is an *analytic proposal*, not a claim that all seven fields are encoded in any one existing standard. It can be linked to the observation and provenance records specified in Section 5 and to the measurement qualifications introduced in Section 4.

The NetCDF Climate and Forecast metadata conventions exemplify why semantic detail matters: standard names, canonical units, spatial and temporal coordinates, and cell methods distinguish the physical quantity and the aggregation it represents [7]. Moreover, term-alignment standards such as SKOS distinguish close matches, exact matches, broader and narrower concepts, and related concepts [70]. A machine-readable link between two labels therefore should preserve its asserted kind and scope; it should not be silently promoted into an assertion of full equivalence. Validation tools such as SHACL may check structural constraints on knowledge graphs, but passing shape validation alone cannot adjudicate a contested scientific interpretation [72].

A federation should report at least three independent results for a proposed exchange: whether the records are structurally valid, whether the claimed mappings are semantically warranted for the particular task, and whether the scientific inference made from translated records remains justified. Failure of any one stage can be consequential even if the other stages succeed. This is particularly important when an AI assistant or automated workflow uses a standards-compliant endpoint to produce an apparently fluent but scientifically misleading synthesis.

6.2 Generative Languages and the Conditional Stability of Scientific Formalisms

Scientific cooperation has often benefited from relatively portable symbols, formal proofs, equations, calibrated units, and agreed procedures for evaluating observations. Yet portability of a particular language is a property of selected representations and practices under specified agreements. Their relational emergence and modes of manifestation do not settle whether the mathematical structures or physical laws to which they refer possess observer-independent ontological reality. The inquiry treats these as analytically distinct questions. A fixed arithmetic proposition can be valid across places and recorded dates under one formal interpretation while the formation, inscription, teaching, and application of that interpretation remain historically situated. A common unit or equation can also mask differences in reference frames, observational supports, causal models, and research purposes. Mathematical invariance, measurement comparability, and identity of generative language must therefore be evaluated separately.

Let a task-specific generalized language of system i be described schematically by

$$\mathcal{L}_i^{(q)} = (\mathcal{D}_i, \mathcal{O}_i, \mathcal{I}_i, \mathcal{P}_i), \quad (21)$$

where $\mathcal{D}_i$ is a locally individuated domain of distinctions or entities, $\mathcal{O}_i$ a family of admissible operations, $\mathcal{I}_i$ procedures of inference and assessment, and $\mathcal{P}_i$ the situated practices and constraints relevant to a declared task q . Equation (21) is an analytic profile, not a claim that a culture or scientific discipline can be exhaustively encoded as four computable sets. These components can evolve and affect one another; some practices may not be captured in a common finite vocabulary. Their use depends on an explicit observational boundary and on who is authorized to characterize the system.

A proposed partial bridge $T_{ij}^{(q)}$ between available manifestations may be assessed only on a declared domain D_q . For a selected observable or action-relevant quantity φ , a local preservation condition takes the form

$$\varphi_j(T_{ij}^{(q)}(m)) = \varphi_i(m) \quad \text{for those } m \in D_q \text{ where both sides are defined.} \quad (22)$$

Equality in Equation (22) is meaningful only after the chosen comparison has been justified; two domains lacking such a bridge need not be metrically commensurable. Even when the displayed equality holds, it does not establish equivalence between the whole generalized languages, the relationally effective structures they describe, or their ultimate ontologies. Approximate and probabilistic counterparts require explicit error or uncertainty profiles. Nor must a shared practical interface preserve all aspects of each system: limited coordination can coexist with significant untranslatable structures.

The nonhuman example clarifies the scope of the proposal. Two GR systems could be responsive to one ontologically shared regularity, while generating distinct inferential and practical structures at the relational register and employing very different media at the manifestational register. A possible observer could manifest a regularity through a patterned chemical process rather than an inscription in arithmetic or calculus. Such an example shows that a particular symbolic form cannot simply be assumed necessary for every conceivable representation. It does not establish that such observers exist, possess a comparable concept of number, or disagree about physical reality. In actual meteorological and oceanographic cooperation, the empirically relevant question is narrower: when instruments, modelers, and institutions use different implementations, which calibrated operations and inferential claims are preserved, under which conditions, with what residual disagreements? Research on culturally situated mathematics and local scientific interlanguages supplies established points of comparison [11, 24].

The infrastructural implication is to maintain distinct records for a language or model's manifest form, the practices and relationships that make it operational, the selected invariants that survive translation, and remaining failures of interpretation. Where distinct manifestations converge on a shared measured regularity, the infrastructure may document that task-specific convergence, while leaving its ontological explanation open. Conversely, lack of a common translation does not prove absence of a shared ontological regularity. Such a dossier cannot settle ultimate ontology. It can, however, help participants decide where joint inference is warranted and where separate descriptions should remain visible. The analytic registers introduced in Section 2.4 thus deepen epistemic interoperability without granting an ontologically privileged position to current scientific notation.

6.3 Contextual Meaning and Measurement Equivalence

Measurement equivalence is sensitive to the target property, the instrument or inference procedure, the observational support, and the question being asked. Two measurements denominated in the same units may concern different depths, spatial footprints, sampling periods, calibration regimes, or environmental states. Conversely, different units may describe the same property when a valid conversion is available. Scientific comparison requires specifying which differences are removable through conversion and which reflect genuinely different observation operations.

Consider surface-temperature records. A point measurement at a particular coastal location, a satellite-derived sea-surface-temperature estimate over an area, and the mean temperature of a modeled grid cell may all be displayed in kelvins. Their common unit permits elementary unit-level operations, but their supports, error structures, spatial averaging, and relation to the physical target may differ substantially. A defensible comparison may require co-location, interpolation, uncertainty propagation, and a model of representativeness. Such operations may introduce additional assumptions that must be recorded, tested, and revisited. Label agreement should therefore be treated as the beginning of an evaluation, rather than its completion.

Let X_A and X_B denote spaces of admissible observation records in two systems, with partial translation $T_{AB} : X_A \rightarrow X_B$. Let Q_A and Q_B be task-specific operations that extract comparable scientific claims from each representation. For a declared domain $D \subseteq \text{dom}(T_{AB})$, one proposed criterion for

task-relative translation adequacy is

$$d_{\mathcal{Y}}(Q_B(T_{AB}(x)), Q_A(x)) \leq \varepsilon(x), \quad x \in D, \quad (23)$$

where $d_{\mathcal{Y}}$ is an explicitly chosen comparison measure and $\varepsilon(x)$ a documented error allowance or uncertainty bound. The expression presupposes that both outputs have first been interpreted in a legitimately common comparison space $\mathcal{Y}$. When such a space cannot be established, Equation (23) is not well formed and cannot be rescued merely by choosing a permissive numerical tolerance. The paper does not assume that a unique or universal $d_{\mathcal{Y}}$ exists.

The domain D matters as much as the error allowance. A unit translation valid for absolute thermodynamic temperature may fail for a coded quality flag, an anomaly with a different reference period, or a domain-specific categorical variable. An aggregation valid over a shared cell may fail if the sensors sample substantially different populations or locations. A responsible translation record must therefore preserve the authorized purpose, the region of validity, the mapping procedure, and the differences it leaves unresolved. Such qualification serves research comparability and is distinct from a license to redistribute the underlying data.

Table 8: Levels of interoperability and the kinds of equivalence they can and cannot establish. All entries refer to task-specific, testable claims rather than universal guarantees.

Level	Candidate evidence	Important unresolved question
Syntactic	Schema conformance, parseable type, accessible interface	Whether corresponding fields have comparable meaning
Terminological	Shared identifier or documented concept mapping	Whether mapped concepts refer to the same scientific target
Metrological	Units, calibration, reference frame, measurement support	Whether measurements are comparable at required accuracy
Inferential	Model assumptions, validation, uncertainty, task-specific tests	Whether translated records support the same scientific claim
Institutional	Authorized use, responsible custodian, correction procedure	Whether the mapping may be used and contested legitimately

Table 8 makes explicit why success at one level does not entail success at the next. In practice these checks may be revisited rather than performed in a strict linear sequence. Scientific disagreement can reveal an error in the initial semantic alignment; new calibration data can invalidate a previously adequate translation; institutional restrictions can change which comparison is authorized. The architecture should retain these histories instead of exposing a single permanent Boolean field labelled “interoperable.”

6.4 Heterogeneous Generative Models

A generative representation describes how a class of observations or inferences may be produced under stated conditions. Depending on the domain, the representation may be a stochastic model, physical field equation, simulation, causal model, statistical observation operator, computational grammar, measurement procedure, or composition of these elements. We use the term *generative* descriptively: it refers to an explicitly characterized procedure or set of admissible operations, and carries no assumption that every generative process is beneficial, autonomous, or fully captured by the selected formalism.

Write a bounded generative description as

$$\mathcal{G}_i = (\mathcal{Z}_i, \mathcal{O}_i, \mathcal{P}_i, \mathcal{I}_i, \mathcal{U}_i), \quad y_i = \mathcal{O}_i(z_i; \theta_i) + \eta_i, \quad (24)$$

where $\mathcal{Z}_i$ is a stipulated space of latent states or structures, $\mathcal{O}_i$ an observation or realization operation, $\mathcal{P}_i$ a family of admissible dynamics or transition relations, $\mathcal{I}_i$ available inference and validation procedures,

and $\mathcal{U}_i$ institutional and technical use conditions. The displayed observation expression applies only when the relevant quantities admit a suitable numeric representation; categorical, logical, and practice-based accounts need not admit that equation. A stochastic interpretation, if used, must specify the distribution of η_i and its dependencies. Neither z_i nor θ_i is presumed directly observable.

This decomposition reveals several different questions commonly obscured by the phrase “model interoperability.” Two models might describe the same target but employ different latent coordinates. They might share output types while presupposing incompatible dynamics. A statistical predictor might reproduce observed values without capturing the mechanisms that would govern a counterfactual intervention. An observation operator may be appropriate within one spatial or institutional boundary and inappropriate outside it. Formal equivalence of output distributions under a restricted dataset therefore does not establish equivalence of physical models, causal claims, or warranted applications.

The design implication is to support *model plurality with recorded dependence* rather than to impose a single global generative grammar. A shared catalogue can preserve model identities, observational and training supports, governing assumptions, validation domains, relevant computational environments, dependence on earlier transformations, and competing explanations. It should represent alternative hypotheses, rejected mappings, and unresolved uncertainty. The aim is not to eliminate disagreement but to make disagreement interpretable and revisable across organizational boundaries.

The same principle applies to AI-assisted discovery of another system’s generative representation. An AI may infer candidate equations, code, constraints, or grammar-like rules from observed outputs. Such reconstruction is an inverse problem: different generative representations may produce observationally indistinguishable outputs within the accessible regime. Its credibility depends on identifiability assumptions, experimental opportunities, domain expertise, independent validation, and adequate treatment of alternatives. Section 4 already established a related observational non-identifiability result; the present framework extends that caution to translation among systems of inference.

6.5 Translation between Observational and Inferential Systems

The proposed federation requires two linked but distinct translation operations. An *observational translation* maps records, support descriptions, units, or calibration status into another usable representation. An *inferential translation* relates the scientific procedures that transform observations into claims. Even where an observational translation succeeds, an inferential translation may fail because the systems employ different assumptions, uncertainty models, or validation criteria.

Let D_i be the authorized and scientifically valid observation domain for system i , and let $F_i : D_i \rightarrow K_i$ map admissible records to claims represented in a system-specific claim space K_i . A comparison may involve partial maps $T_{AB} : D_A \rightarrow D_B$ and $S_{AB} : K_A \rightarrow K_B$. A bounded translation diagram is *commutative for the selected claim and domain* if

$$S_{AB} \circ F_A(x) = F_B \circ T_{AB}(x), \quad x \in D \subseteq D_A, \quad (25)$$

whenever both sides are defined. Equation (25) is a proposed adequacy test, not an assertion that such maps always exist. Equality may be replaced by an explicitly justified task-specific tolerance as in Equation (23), and different applications may require different tests. A successful diagram for one property does not authorize its reuse for another scientific claim.

The notation also separates representations from the persons and institutions entitled to assert or challenge them. The custodian of F_A may contest the premise that its procedures are captured accurately by the translation. Affected contributors may restrict uses of some records. Independent reviewers may discover that a claimed correspondence held only under a narrow environmental regime. Each of these events can alter the admissible domain D , modify a proposed mapping, or trigger a revised statement of uncertainty.

The following workflow gives this mathematical proposal a feasible investigative form. A requesting system states its research question, authorized purpose, target phenomenon, and required precision. Participating custodians supply available observational descriptions, relevant versions, model assumptions, and restrictions. Candidate alignments are generated by a documented combination of vocabulary

matching, measurement transformations, scientific modeling, and human review. The proposed mapping is tested on independent cases designed to reveal known ambiguities or systematic disagreement. Its limitations, contested interpretations, and failure modes are returned alongside any translated claims. The output is a versioned scientific artifact, not an unqualified assertion that the systems have become equivalent.

A design requirement is that mappings are not silently composed merely because intermediate systems consider them acceptable. If T_{AB} and T_{BC} are each adequate under different supports or purposes, $T_{BC} \circ T_{AB}$ need not be adequate for the original task. Domain restrictions, accumulated uncertainty, governance constraints, and loss of contextual information must be checked again. The relation between this problem and ontology alignment is illustrated by SKOS’s explicit distinction between close matches and exact matches [70]. Traceable mappings also require provenance over mapping decisions, model revisions, and reviewing responsibilities, which may be represented with the activities, entities, and agents of PROV-O [71].

The cross-layer review contract in Equation (18) provides the engineering counterpart of this epistemic distinction. A task-preservation claim should be linked to the scientific reviewer and test domain, while a proposed use of its resulting artifact must carry a separately sourced assessment of competent permission. Two mediations that are individually supported need not produce a supported composed mediation: if the intermediate representation suppresses a distinction required by the destination task, no subsequent fluent mapping can restore that distinction without additional evidence. An implementation should therefore check the composition against the destination inquiry and disclose unresolved or refuted preservation claims, as further developed in Section 10.4.

6.6 Loss, Noninvertibility, and Untranslatable Structures

Translation commonly discards information. Spatial aggregation eliminates subgrid differences, temporal averaging removes part of a trajectory, quantization obscures finer measurements, and category alignment may suppress distinctions that matter to a particular community or scientific field. Some losses can be quantified. Others are structural: the receiving representation lacks a type, a relation, or an institutional category necessary to express what was present in the source. The federation should recognize both rather than promise lossless conversion between all knowledge systems.

Proposition 6.1 (Noninvertibility of aggregation for an unpreserved claim). *Let X be a set of source histories, $T : X \rightarrow Y$ a translation or aggregation, and $q : X \rightarrow Q$ a target scientific quantity. If there exist $x_1, x_2 \in X$ such that $T(x_1) = T(x_2)$ but $q(x_1) \neq q(x_2)$, then no function $\hat{q} : Y \rightarrow Q$ can reconstruct q exactly from the translated record for every $x \in X$.*

Proof. If such a $\hat{q}$ existed, then $q(x_1) = \hat{q}(T(x_1)) = \hat{q}(T(x_2)) = q(x_2)$, contradicting the assumed difference. The obstruction is informational and does not depend on the computational power of the proposed reconstruction system. $\square$

An elementary example uses two hypothetical two-day temperature histories in degrees Celsius: $x_1 = (10, 30)$ and $x_2 = (20, 20)$. Both have a two-day mean of 20, while their respective maxima are 30 and 20. If T retains only the mean, the maximum cannot be recovered uniquely. Proposition 6.1 is deliberately elementary; its function is to demonstrate a strict logical boundary on claims of universal AI-assisted knowledge reconstruction. Additional measurements, assumptions, or experimental intervention could distinguish the histories, but they would supply *new information*, not reverse an information-losing map by computation alone.

An analogous obstacle may occur when two institutional histories generate the same publicly visible record despite different consent conditions or contribution relationships. Metadata that omit these relations cannot reconstruct them from an identifier alone. Equally, two models with the same output on a finite evaluation sample can disagree under intervention or distribution shift. Whether a practical translation is adequate thus depends on which distinctions an investigation needs to preserve and what independent evidence is available.

A proposed translation manifest can report a vector of losses

$$\mathcal{L}(T \mid D) = (L_{\text{measurement}}, L_{\text{support}}, L_{\text{model}}, L_{\text{provenance}}, L_{\text{rights}}), \quad (26)$$

where each entry refers to a separately defined test or a documented qualitative limitation. Equation (26) does not assert that these quantities share units, are additively commensurable, or can legitimately be collapsed into a single quality score. For some cases a dimension may be unknown or fundamentally contested. Preserving that result is more informative than substituting an arbitrary numeric value.

6.7 Epistemic Bias and Contested Interpretation

A federation can reproduce epistemic inequality even when it improves technical interoperability. Vocabulary selection may privilege highly resourced observing communities; machine translation may flatten distinctions made in local languages; model-selection procedures may reward systems with the largest training datasets; and evaluation may ignore observation gaps in less intensively monitored regions. Such issues concern the conditions under which some claims become legible and authoritative, not simply the accuracy of data formatting. Related discussions of situated knowledge, collective criticism, and epistemic injustice provide relevant philosophical resources [32, 46, 23].

An adequate design must make semantic decisions contestable. Contributors should be able to propose alternative labels, report invalid mappings, identify incomplete provenance, contest the applicability of inferred models, and document harms arising from automated reinterpretation. There may be legitimate reasons to restrict a dataset’s location details, access conditions, or downstream inference, particularly in sensitive ecological and community contexts. Publicly discoverable metadata should not be presumed to entail rights to retrieve, analyze, or republish the corresponding data. Nor should a successful machine translation confer the authority to speak on behalf of the contributing community.

Disagreement is not always evidence of a system defect. Two observation systems may use different scientifically defensible measurement conventions or may target different aspects of an incompletely understood phenomenon. A federation should support parallel interpretations, traceable reasons for divergence, and the possibility of unresolved disagreement. When a discrepancy is an identifiable error, clear procedures for correction and notification are needed. When the discrepancy reflects competing theoretical commitments, the relevant assumptions and the evidence available to assess them should remain visible rather than being suppressed in a false consensus.

These requirements imply a provisional *translation review record*: the originating and receiving representations; the stated target and use; the mapping type and supporting assumptions; observation and model versions; validation evidence and uncertainty; information known to be lost; access and contribution conditions; dissenting interpretations; and the responsible route for correction or appeal. This record extends the provenance and governance interfaces described in Section 5. It does not presume that every dispute can be resolved computationally or that all jurisdictions recognize the same procedures for remedy.

A recurrent source of conflict can be represented without supposing that one participant’s understanding is simply false. For an observation record e , interpretive system G_i , situated history h_i , and research purpose q , let

$$I_i(e \mid G_i, h_i, q) \in \mathcal{K}_i, \quad I_i(e \mid G_i, h_i, q) \not\equiv I_j(e \mid G_j, h_j, q), \quad (27)$$

where $\not\equiv$ indicates that the outputs have not been established equivalent for the specified inquiry; it is not a claim that the two knowledge spaces admit a canonical inequality. The same temperature, species-occurrence, or satellite observation might be used to support different judgments about scientific reliability, evidentiary sufficiency, or acceptable disclosure. Translation systems that preserve numerical fields but omit the generating histories and evaluative purposes may inadvertently make such divergences invisible.

The ensuing dispute can be both epistemic and normative. Participants may disagree about what an observation warrants, about which uses of a warranted result are acceptable, or about who has standing to decide. These are analytically separable questions; better sensor accuracy alone may not decide a conflict

about endangered-species disclosure, while a negotiated access arrangement does not establish which ecological model is correct. Norm-contestation research already investigates divergent meanings and practices in international ordering [60]; the further question here is how an observation federation could preserve such differences in records and support a limited, reviewable cooperative procedure. Section 8.7 specifies that procedural interface.

The argument of this section can be summarized carefully. Shared schemas, vocabularies, and model registries provide valuable conditions for global scientific cooperation. They do not create a universal epistemic standpoint. The feasibility of translation depends on the particular phenomena, observations, model assumptions, research purposes, institutional rights, and revisions at issue. Section 7 will examine how these conditional mappings interact with data assimilation, model uncertainty, and independent validation. The companion governance paper will address the broader formation and repair of trust and recognized authority in greater detail.

7 Data Assimilation, Model Plurality, and Uncertainty

Sections 4–6 distinguished situated observations, federated access, and task-relative translation of heterogeneous representations. The present section examines the inferential layer through which observations are brought into contact with dynamical models and through which an infrastructure might make multiple, partly incompatible estimates scientifically usable. Its objective is to specify what an assimilation result means, which uncertainty assumptions authorize particular calculations, and how competing products may be compared without regarding one numerical field as the uniquely warranted image of the world. The motivating cases are atmospheric and oceanographic data assimilation, but the proposed requirements concern a broader federation of knowledge-generating systems. In particular, global geographic coverage does not imply complete observational coverage, and dense numerical outputs do not imply that an unobserved physical quantity has become directly measured.

Operational data assimilation is an established scientific and engineering field. The European Centre for Medium-Range Weather Forecasts (ECMWF) describes its analyses as combinations of observations and model estimates, including atmospheric and ocean–sea-ice components [16]. Climate reanalysis similarly reconstructs earlier environmental states with a model and archived observations; it does not reverse history to generate the measurements that were absent [15, 33]. The formal constructions below use familiar filtering, variational, and ensemble ideas as *illustrative methods*, not as a claim that a general-purpose federated assimilation service already exists. A federation would additionally need methods for deciding which transformations are scientifically appropriate, which records may lawfully be used, and how disagreements and corrections are preserved.

7.1 From Observation to Inference

A scientific observation ordinarily registers a signal or reported quantity under an identified protocol, support, and calibration context. An analysis instead estimates an environmental state or other target in light of observations, model dynamics, prior information, and a declared inferential scheme. For a selected modelling problem, let $x \in \mathcal{X}$ denote a latent state, $y \in \mathcal{Y}$ an observation vector, $H : \mathcal{X} \rightarrow \mathcal{Y}$ an observation operator, and $M : \mathcal{X} \rightarrow \mathcal{X}'$ a model continuation operator. A local linearized observation representation is

$$y = H(x) + b + \varepsilon, \quad x^b = \mathcal{M}(x_{\text{ref}}) + \eta, \quad (28)$$

where x^b is a background or prior estimate, b a possibly structured measurement or representation bias, ε observational error, and η an aggregate discrepancy in the stated background model. The two terms b and η have deliberately different interpretations. Neither is safely assumed independent, stationary, Gaussian, or fully known without domain-specific evidence. The reference x_{ref} and map $\mathcal{M}$ refer to a specified model history and need not imply a universal clock or a privileged ordering of all events.

In a Bayesian formulation, the analysis is a conditional distribution rather than necessarily a single best point estimate:

$$p(x | y, m, \iota) \propto p(y | x, m, \iota) p(x | m, \iota), \quad (29)$$

where m denotes a chosen model family and ι bundles the observation-selection rule, calibration versions, inferential protocol, and other declared background information. Equation (29) is valid as a representation when an appropriate probabilistic model has been specified; it does not guarantee that its likelihood, prior, or model family is correct. A posterior mean, credible interval, or forecast is therefore conditional on m and ι . Different admissible analyses may appropriately disagree. The mere delivery of an analysis by a prestigious institution does not turn it into unconditioned observational truth.

These distinctions are made concrete in reanalysis. ERA5 combines observations and an atmospheric model to produce spatially extensive environmental estimates, with a separate ensemble-based uncertainty characterization; reported limitations include changes in the historical observing system and systematic biases not fully represented by ensemble spread [33, 9]. For a proposed global architecture, the minimum traceable unit should therefore link each derived analysis version to its input observation versions, model identity, assimilation-window convention, observation operator, uncertainty assumptions, and record of diagnostics. A grid cell with no local measurements can still have a model-based estimate, but the infrastructure should not describe it as if that value were a newly observed measurement.

7.2 Assimilation Assumptions and Model Error

The algebra of common assimilation schemes makes the dependence on assumptions especially visible. In a finite-dimensional linear Gaussian illustration, suppose x^b has error covariance B , the observation error has covariance R , and the linear observation operator is H . If B and R are positive definite, the posterior mean update and gain are

$$K = BH^T(HBH^T + R)^{-1}, \quad x^a = x^b + K(y - Hx^b), \quad (30)$$

with the conventional idealized posterior covariance

$$P^a = (I - KH)B = (B^{-1} + H^T R^{-1} H)^{-1}. \quad (31)$$

The equalities in Equation (31) hold in the stated linear Gaussian model with the specified independence and covariance assumptions; they are not generic identities for nonlinear or misspecified assimilation. Variational methods and ensemble Kalman filtering extend parts of this framework under distinct approximations and numerical conditions [22, 16].

A model for an evolving environmental state may instead use $x_{j+1} = M_j(x_j) + \eta_j$ along a model-specific ordered history. Its process-error law, observation operators, and calibration changes must be declared. In an ocean setting, altimetry, hydrographic profiles, moorings, and satellite retrievals can constrain different state components and spatial scales. Model dynamics help connect these incomplete observations, but can also propagate incorrect parameterizations or poorly represented boundary processes. More measurements do not necessarily cure structural model error: high-frequency satellite coverage, for example, may offer limited information about a poorly observed deep-ocean process without an adequate physical observation operator.

An important distinction is between *forecast spread*, *analysis uncertainty*, and *model adequacy*. A narrow distribution of predictions under one set of perturbations can coexist with systematic bias shared by every ensemble member. ECMWF’s guidance for ERA5 explicitly cautions against treating its ensemble-of-data-assimilations uncertainty values as direct numerical measures of all reanalysis errors [9]. An infrastructure consequently should preserve the versioned source of B , R , error correlations, treatment of bias, and the methods used for quality control. It should also record excluded observations and reasons for exclusion. The ability to recompute an algorithm’s nominal covariance is valuable, but cannot replace independent testing of whether that covariance is calibrated for the scientific task.

A simple counterexample demonstrates why the observation-error structure matters. Let a scalar state $x \sim N(0, 1)$ be measured twice, but let $y_1 = y_2 = x + \varepsilon$ with *the same* error $\varepsilon \sim N(0, 1)$ in both records.

There is one independent item of information. The correct posterior variance is $1/2$. If the two records are wrongly treated as independent unit-variance measurements, the nominal posterior variance becomes $1/3$. The second value appears more precise but is unsupported by the data-generating arrangement. This is relevant when a federation inadvertently ingests replicated observations, overlapping satellite retrievals, or derived products that reuse the same raw inputs. Shared provenance is therefore part of error specification, not merely bibliographic decoration.

7.3 Model Ensembles and Structural Disagreement

A federation may connect models constructed for different targets, scales, disciplines, and policy questions. Some differences are parametric: multiple plausible parameter vectors inhabit one common mathematical representation. Others are structural: models distinguish different state variables, governing processes, constitutive relations, or acceptable boundary conditions. A single ensemble generated by perturbing one model family primarily samples variability within the assumptions used to construct that ensemble. It does not exhaust the space of plausible physical interpretations.

Let $\mathcal{M} = \{m_1, \dots, m_J\}$ be a declared finite set of model hypotheses, and let $q(x)$ be a scientifically specified derived quantity. Within a coherent probabilistic model-averaging framework, if the models have justified weights $w_j \geq 0$ with $\sum_j w_j = 1$, the law of total variance yields

$$\text{Var}(q \mid y) = \sum_{j=1}^J w_j \text{Var}(q \mid y, m_j) + \sum_{j=1}^J w_j (\mu_j - \bar{\mu})^2, \quad \bar{\mu} = \sum_j w_j \mu_j, \quad (32)$$

where $\mu_j = \mathbb{E}[q \mid y, m_j]$. The first term in Equation (32) describes within-model uncertainty and the second describes disagreement among conditional model means, under the stipulated mixture. Assigning w_j is itself a substantive scientific assumption. If no defensible common probabilistic weighting exists, an ensemble of competing models should instead report distinct claims, uncertainties, domains, and disagreements without a synthetic posterior pretending to resolve them.

Proposition 7.1 (Observational agreement does not identify continuation). *Equality of observable values for a finite record does not, without further assumptions, imply equality of model-implied continuations of those values.*

Proof. Consider states $x = (a, b) \in \mathbb{R}^2$ with observation operator $H(a, b) = a$. Let two deterministic continuation maps be $M_1(a, b) = (a, b)$ and $M_2(a, b) = (a + b, b)$. For the same initial state $x = (0, 1)$, both models produce the observed value $H(x) = 0$. Their continued predictions differ: $H(M_1(x)) = 0$ and $H(M_2(x)) = 1$. Thus identical fit to the selected initial observation does not determine a common continuation. The example establishes logical underdetermination, not a claim that either model is physically appropriate for a particular Earth-system process. $\square$

Proposition 7.1 matters whenever an AI mediator or central registry is tempted to replace several context-specific analyses with a single “best” model on the basis of agreement at observed locations. Matching data may underdetermine models for extrapolation, counterfactual inquiry, rare extremes, or interventions. Moreover, assigning a single ensemble weight to an institutional disagreement about a hazardous coastal event could hide the reasons that affected communities, regional scientists, and global modelling centres interpret the relevant evidence differently. Model plurality thus has epistemic value insofar as its distinctions are preserved and can be tested; plurality is not itself a guarantee of accuracy or fairness.

7.4 Propagation of Measurement and Translation Uncertainty

Global observation data often pass through multiple conversions before entering a common analysis: coordinate conversion, unit handling, spatial aggregation, bias adjustment, interpolation, semantic mapping, and model-specific observation operators. The operations may change information content and

error dependence. We therefore propose that a federation document the uncertainty consequences of each transformation rather than attach one undifferentiated quality score to the final product.

For a differentiable translation T evaluated locally at a signal mean $\bar{y}$, a first-order covariance propagation approximation takes the form

$$\Sigma_{T(y)} \approx J_T(\bar{y}) \Sigma_y J_T(\bar{y})^\top + \Sigma_{\text{map}}, \quad (33)$$

where J_T is the Jacobian, Σ_y the covariance of the input record, and Σ_{map} an explicitly modelled additional mapping-error covariance. The additive form in Equation (33) assumes that any cross-covariance between input error and mapping error is negligible or already handled. Where that assumption fails, the corresponding cross terms are required. Nonlinear translations, bounded quantities, categorical observations, rare events, and non-Gaussian distributions may need sample-based propagation or carefully justified alternative uncertainty representations.

No covariance propagation repairs information lost through an inappropriate aggregation. Consider a deep-water profile and a coastal surface measurement that a portal converts to the same standardized “temperature” descriptor. Their uncertainty intervals might be individually narrow, while the comparison remains physically ill-posed because the target supports differ. Similarly, an hourly mean cannot reveal the maximum of a short extreme event without additional assumptions, as discussed in Section 6. More precise metadata can make the invalid comparison detectable; it cannot make the comparison valid.

It is therefore useful to record an *uncertainty and transformation dossier* D_q for every significant derived scientific claim q :

$$D_q = (\mathcal{O}_q, \mathcal{T}_q, \mathcal{M}_q, \mathcal{U}_q, \mathcal{V}_q, \mathcal{A}_q, \mathcal{C}_q), \quad (34)$$

where $\mathcal{O}_q$ identifies observation versions and sampling support; $\mathcal{T}_q$ the translation chain; $\mathcal{M}_q$ models and parameters; $\mathcal{U}_q$ quantified and unquantified uncertainties; $\mathcal{V}_q$ validation evidence; $\mathcal{A}_q$ access and authority rules; and $\mathcal{C}_q$ correction and contestation records. Equation (34) is an exploratory research schema, not a representation implemented by one of the existing standards or programmes. Some components may be held in federated restricted stores. Discoverability does not entail a right to publish sensitive observation coordinates or protected source data.

Table 9: Sources of uncertainty and their distinct consequences in an interoperable observation and assimilation workflow.

Source	Illustrative origin	Required documentation or check
Measurement	Sensor noise, drift, calibration change	Calibration version, reference evidence, observation-error model
Sampling support	Missing depths, uneven regions, changes in coverage	Spatial/temporal support, missingness mechanism, coverage diagnostics
Translation	Aggregation, interpolation, coordinate conversion	Valid domain, method, induced loss, cross-covariances
Assimilation	Background covariance, observation operator, bias	Algorithm/version, tuning, residual diagnostics, rejected observations
Model structure	Missing processes, boundary conditions, regime change	Parallel models, independent evaluation, domain-of-validity statement
Institutional use	Restricted inputs, authority or interpretive disagreement	Authorized use, contestation path, responsible custodian

Table 9 distinguishes scientific error sources from institutional use conditions. The latter cannot be converted into a covariance simply to obtain a numerical optimization problem. Nonetheless, institutional control can indirectly change scientific uncertainty by determining which observations, computational tools, and local expertise can participate in an analysis. The result is a relational dependence

between epistemic capacity and governance, not a claim that every social asymmetry has a measurable probabilistic effect.

7.5 Outlier Detection, Reanalysis, and Independent Checking

A proposed knowledge infrastructure must permit both routine automated quality control and human scientific disagreement about unusual records. An innovation $d = y - H(x^b)$ may be large because an observation is defective, but also because the model misses an extreme event or because the comparison uses an incorrect observation operator. Rejecting unusual values indiscriminately can therefore suppress the very phenomena that a global observing network ought to discover. Conversely, admitting every unexplained outlier into an analysis can damage model estimates. An appropriate protocol records quality-control decisions, alternative explanations, uncertainty assumptions, and routes for revision.

For an observation-selection indicator $s_i \in \{0, 1\}$, the analysis obtained from a selected dataset may be written as $\mathcal{A}(\{(y_i, s_i)\}_{i \in I}, m, \iota)$. The choice of s_i is part of the inferential history, not merely a pre-processing operation that can be forgotten. Scientific reporting should distinguish whether observations were rejected because of instrument flags, internal consistency tests, representativeness doubts, or policy restrictions. The high-confidence appearance of an assimilated global map does not justify discarding the records that explain how its observational support was assembled.

Reanalysis provides an especially informative example of historical revision. ECMWF describes reanalyses as combinations of archived observations and forecast models using a consistent analysis system, while noting that inputs are digitized, reprocessed, and quality controlled across production cycles [15, 33]. A revised analysis can change an estimate of a previously observed condition without changing the condition itself. Revisions therefore do not literally alter the represented physical history; they alter what can be warranted about it given changed data, models, and methods. The relevant relational history comprises versions, dependencies, and scientific judgments, with no requirement to treat a version identifier as a universal temporal coordinate for all domains.

Independent checking should not be confused with comparing an analysis to observations already used to construct it. The latter can provide useful diagnostic residuals, but cannot by itself demonstrate out-of-sample predictive validity. A federation should document whether evaluation uses withheld observations, independently calibrated instruments, separately developed models, independent regional expertise, or observations obtained through subsequent field campaigns. Where data are sparse, uncertainties about validation representativeness should remain explicit. Separate institutions can reproduce the same algorithmic result and still share the same observational blind spot, bias, or code dependency.

A particularly valuable policy is to make *correction without erasure* possible. Superseded observational records and analyses should be linked to revised records and explanatory notices, subject to lawful retention, privacy, and community data governance. An invalid scientific claim should no longer be presented as authoritative, yet a sufficiently documented history of its production may be necessary for understanding why a public warning, environmental decision, or research conclusion was made. An audit log makes historical inquiry possible; a formal mechanism for correction, appeal, and remedy makes such inquiry institutionally meaningful.

7.6 Validation across Institutions and Domains

Evaluation across a global knowledge federation poses two linked questions: whether an analytical procedure behaves as claimed within its stated scientific domain, and whether its outputs remain responsibly interpretable when used across institutional boundaries. A meteorological assimilation tested against withheld radiosonde observations is not thereby validated for deep-ocean ecology. An ocean-reanalysis uncertainty model calibrated for basin-scale temperature patterns may not warrant a coastal habitat intervention. Even within the same field, a hindcast can show good average skill while missing rare events of exceptional public importance. Scientific validation is thus *task-relative*, *region-sensitive*, and *historically revisable*.

Table 10: Proposed cross-institution validation dimensions; these are research criteria rather than measured performance of an existing federation.

Dimension	Candidate test	Principal failure mode to examine
Observation independence	Withheld sites, duplicate detection, provenance overlap audit	Hidden reuse of the same measurements
Model plurality	Compare structurally distinct analyses on common tasks	Agreement produced by shared assumptions
Uncertainty calibration	Coverage, reliability, spatially stratified diagnostics	Underdispersion, systematic bias, local failures
Domain transfer	New basin, sensor type, regime, or event category	Invalid extrapolation and mismatched support
Historical revision	Reproduce specified versions and trace corrections	Undocumented changes in input or processing
Institutional accountability	Source challenge, access dispute, independent response	Inability to correct or contest consequential use

The dimensions in Table 10 can be converted into pre-registered empirical protocols, but should not be combined into one global performance index without an explicit justification of their normative and scientific weighting. Climate monitoring, disaster warning, deep-sea habitat assessment, and protected community research may have different tolerances for uncertainty, confidentiality, latency, and harmful error. Even the common term “independent validation” needs qualification when institutions depend on the same satellite measurements or shared postprocessing libraries.

One minimal validation condition can be expressed as follows. For a declared scientific target q , deployment domain D , analysis procedure $\mathcal{A}$, and institutionally approved evaluation protocol Π , define the evidence record

$$\text{Valid}_{\Pi}(\mathcal{A}, q, D) := (\mathcal{E}_{\text{held out}}, \mathcal{E}_{\text{robust}}, \mathcal{E}_{\text{cal}}, \mathcal{E}_{\text{audit}}; \Pi). \quad (35)$$

The symbol Valid_{Π} in Equation (35) denotes an evaluable *bundle of evidence*, not an unconditional Boolean theorem that an analysis is true. The components concern independent test data, sensitivity and model-robustness tests, uncertainty calibration, and a traceable institutional review. They may yield mixed or disputed assessments; the protocol must also record who can challenge and amend them. A future federation can support such assessments by preserving interoperable records, shared test specifications, and authorized verification services. It cannot make their scientific and political judgments disappear by standardizing their file formats.

The resulting research agenda is specific. Scientific agencies and communities would need to test model ensembles across heterogeneous observation supports, quantify the additional uncertainty from translation, report data and code dependencies behind apparently independent assessments, and evaluate whether correction and dispute processes function in practice. Existing meteorological and oceanographic systems offer substantial operational experience, including the documented limitations of ERA5 ensemble-based uncertainty estimates [9, 16, 10]. The proposed global infrastructure extends the accountability requirements beyond current examples and remains a design hypothesis requiring empirical and institutional assessment. Section 8 will turn to trust formation, credible commitments, and relational governance; its argument must not presuppose that successful assimilation has already established cooperation, consent, or shared authority.

8 Trust, Credible Cooperation, and Relational Governance

The preceding sections specified how a federation may preserve situated observations, expose their provenance, coordinate access across heterogeneous institutions, translate representations under declared conditions, and evaluate competing analyses. None of those capabilities entails that a participant has reason to entrust another with an observation, accept its interpretation, consent to a particular use, or believe that a consequential error will be corrected. The present section addresses this additional institutional problem. Its object is the *formation and maintenance of credible cooperation* in global observation and knowledge infrastructures: the history of commitments and their performance, the allocation of authority, the ability to challenge decisions, and the practical accessibility of correction and remedy. It is a preliminary governance discussion for an infrastructure design; a fuller theory of generative trust and transnational institutions is reserved for the companion investigation described in Section 8.8.

The starting point is neither an assumption of generalized mistrust nor a presumption that scientific collaboration eliminates conflicting interests. Institutions may cooperate in disaster warning while retaining different legal duties, interpretation practices, security constraints, financial resources, and political priorities. Existing meteorological exchange is informative precisely because it operates through an assemblage of national contributors and agreed services. WIS2, for example, distinguishes publishing nodes, Global Brokers, discovery catalogues, caches, and monitoring functions; the WMO data policy also distinguishes unrestricted core data from recommended data that may involve access conditions [66, 63]. These arrangements establish operational roles and shared protocols, not a universal theory of justified reliance. More general work on collective-action institutions and polycentric governance provides an analytical vocabulary for investigating situated rule formation, monitoring, and trust without equating one legal form or one technical architecture with legitimate cooperation [51, 52].

8.1 Trust Formation and Institutional History

Trust is not treated here as a fixed attribute of an organization, a numerical reputation score, or an inherent property of a data product. For a specified relation, it concerns an actor’s warranted or unwarranted practical reliance on another actor or institutional arrangement under uncertainty, exposure, and partly divergent interests. A meteorological service may rely on a partner’s data-delivery schedule while contesting its calibration procedures; it may trust a national observatory’s measurements for one scientific target yet require independent verification for an emergency intervention. Trust relations are therefore object-specific, context-sensitive, and revisable.

A useful provisional representation takes a documented relational history h containing events of commitment, performance, failure, explanation, correction, and institutional response. For participants i and j , an entrusted activity or claim q , and an explicitly declared decision context ι , write

$$\mathcal{T}_{i \rightsquigarrow j}(q \mid h, \iota) := (E_{ijq}, V_{ijq}, R_{ijq}, U_{ijq}; h, \iota), \quad (36)$$

where E records expectations and reliance, V evidence concerning previous performance and verifiability, R the availability and operation of review or remedy, and U unresolved uncertainty or disagreement. Equation (36) is a *descriptive dossier*, not a trust function with an objective scalar value. Even complete records of earlier performance cannot logically guarantee a future action. Equally, an absence of past disputes may reflect successful cooperation, limited participation, or inaccessible channels for complaint. The distinction matters when a network equates quiet operation with legitimacy.

Histories may be partially ordered by documented dependence: an amended dataset relies on an earlier observation and a later calibration result; a governance decision may depend on an independent scientific review. No universal external clock is required for every organizational process to share one temporal metric. Observable ordering and causal attribution nevertheless remain substantive empirical claims. The federation should be able to record when particular commitments were declared, which evidence was available to whom, and which events were prerequisites for authorization or withdrawal. An institutional record that merely displays the latest endorsed state suppresses the circumstances under which reliance became reasonable, unreasonable, or contested.

Trustworthiness, reliance, and trust should accordingly be distinguished. An organization may behave competently without being trusted because its counterpart lacks access to the relevant evidence. It may be trusted despite failures because affected participants have few alternatives. Institutions may also extend reliance through enforceable contracts, redundant observations, independent oversight, or jointly established protocols without requiring a deep personal relationship. These substitutions and complementarities are part of the research problem. The value of collective monitoring and appropriately matched polycentric rules for cooperation is documented in Ostrom’s work; direct transplantation to a global scientific federation remains a hypothesis requiring comparative assessment [51, 52].

8.2 Verification, Authentication, and Practical Reliance

Authentication establishes a claimed source or identity under a chosen trust anchor. A cryptographic signature can support the verification that specified bytes were signed by a holder of a corresponding key. A checksum can help detect an altered file. A provenance graph can identify records and processes presented as antecedents of a result. These are important properties, but each establishes a different proposition from whether an observation is accurate, whether a model is adequate, whether its use is permitted, or whether a decision is justifiable. This difference applies even when every node follows the same metadata standard.

Consider a signed temperature series produced from an incorrectly calibrated sensor. The signature can be valid, the data format correct, the submitting organization duly registered, and the reported quantity systematically biased. Conversely, a scientifically valuable observation from an independent community contributor may lack recognized platform credentials. The first can pass a technical authenticity check while failing a scientific task; the second can deserve substantive consideration while requiring provenance and consent procedures appropriate to its use. A federation cannot coherently treat either authenticity or institutional standing as a universal proxy for scientific quality.

The proposed evaluation therefore separates at least five questions: source authentication, record integrity, scientific validity for a specified purpose, permission to access and reuse, and accountable institutional endorsement. For a data item d , task q , user or requesting institution u , and governance context g , define a multi-part assessment as

$$\mathcal{V}(d, q, u, g) = (v_{\text{auth}}, v_{\text{int}}, v_{\text{sci}}(q), v_{\text{permit}}(u, g), v_{\text{review}}(g)). \quad (37)$$

The components in Equation (37) may be unknown, contested, or supported by incompatible evidence; they are deliberately not multiplied into a single certificate. For example, v_{sci} cannot be inferred from $v_{\text{auth}} = 1$. These conceptual dimensions are design requirements rather than existing standardized global fields.

WIS2 illustrates both the possibilities and the boundary of technical coordination. Its Global Discovery Catalogue facilitates discovery, Global Brokers disseminate notifications, and Global Caches make relevant core data available, while data classified as recommended may require retrieval from the publisher under stated access arrangements [66, 64]. System monitoring and identification of service operators can support operational reliability. Such facilities do not independently adjudicate conflicts over underlying measurement practice, environmental risk, or the appropriateness of downstream AI-generated interpretations. A useful federation should accordingly permit provenance inspection and scientific challenge by actors who are institutionally affected but not necessarily the operators of its main registry.

The dimensions of Table 11 jointly inform responsible reliance, but their weights cannot be specified independently of the decision context. An emergency early-warning service and an archival research catalogue may have sharply different latency, error, confidentiality, and response requirements. A more capable AI classifier can reduce the cost of preliminary anomaly detection; it does not acquire by that fact the authority to overrule a community’s legitimate restrictions or to discharge an institution’s legal responsibilities.

Table 11: Distinct grounds of practical reliance in a federated observation service. The proposed checks are illustrative and do not certify existing systems.

Question	Evidence or candidate check	Important remaining uncertainty
Who provided the record?	Signing keys, accountable identifiers, registry status	Credential ownership, compromised keys, recognition barriers
Was it preserved?	Hashes, version records, dependency links	Omitted inputs and incomplete histories
Does it support the claim?	Calibration, withheld validation, model diagnostics	Bias, task mismatch, structural inadequacy
May it be used?	Consent, data-use conditions, legal review	Conflicting laws, community authority, secondary use
Who can correct a mistake?	Appeal routes, notices, remedy and version links	Unequal procedural access, delay, ineffective enforcement

8.3 Reciprocal Commitments and Obligations

Cooperation becomes more credible when its participants can identify what has been promised, to whom, under which conditions, with what evidence of performance, and through which processes breaches can be addressed. A shared data standard is one element of such an arrangement. Other elements may include calibration obligations, reasonable delivery intervals, revision notices, confidentiality safeguards, acknowledgement of contributors, reciprocal access to computational resources, support for low-capacity nodes, or independent methods for verifying compliance. These commitments should be distinguished from goodwill or a presumption of permanent consensus.

Let a proposed interinstitutional commitment be a typed record

$$c = (i, j, a, \chi, e, \omega, r, \kappa), \quad (38)$$

where i undertakes an activity a toward beneficiary or counterparty j under applicability conditions χ ; e specifies what would count as evidence of performance, ω a review and escalation process, r available corrections or remedies, and κ the authority and jurisdictional scope under which the record is recognized. Equation (38) is a proposed representational grammar. It is not a claim that any existing international treaty or WMO service uses these fields, nor that every institutional commitment is legally enforceable.

Credibility depends in part on whether these fields refer to *practically available* institutions. A nominal right to challenge an erroneous product is weak when an affected institution cannot inspect the relevant metadata, submit evidence in an accessible language, access a competent reviewer, or receive any response before the decision produces irreversible harm. Similarly, reciprocal publication duties can be asymmetric where only some participants possess reliable energy, storage, network access, or trained personnel. Reciprocity should therefore be analyzed as a relational and institutional arrangement rather than as the equality of nominal obligations. Historical contributions and benefit flows may be heterogeneous and cannot all be reduced to bytes transferred or financial payments.

A commitment can also fail even where verification is technically possible. One participating node may meet a publication deadline while omitting a known calibration caveat; another may openly publish sensitive habitat locations in violation of a legitimate protection policy; a third may continue to cite a superseded forecast after receiving notice of correction. The ledger of fulfilled data-delivery events would be insufficient to evaluate these cases. Credibility requires the ability to compare the actual act with its scientifically and institutionally specified purpose, not merely to check whether a message arrived.

One potential design is to associate each commitment c with admissible evidence bundles $\text{Ev}(c)$ and a publicly specified, but access-controlled when appropriate, review procedure $\text{Rev}(c)$. Different commitments can legitimately require different observers and verification methods. An external auditor may be authorized to inspect metadata while a community authority alone can approve reuse of a sensitive resource record. This preserves different dimensions of accountability while maintaining an interoperable

interface. The UNESCO Recommendation on Open Science recognizes both international cooperation and justified data-access restrictions, including protections for sensitive indigenous knowledge and endangered species; it is a normative international recommendation rather than an implementation of the proposed commitment grammar [57].

8.4 Disagreement, Contestability, and Repair

Scientific disagreement need not be a sign of institutional failure. Different observing supports, calibration histories, disciplinary concepts, and model assumptions can yield defensible, competing interpretations. Effective cooperation requires the capacity to distinguish such disagreement from record manipulation, unauthorized use, negligence, and failures of procedural participation. A governance arrangement that treats contestation as a threat to trust may inadvertently suppress the evidence required to make cooperation trustworthy.

The proposed infrastructure should therefore support at least three kinds of challenge. A *record challenge* concerns the integrity, calibration, provenance, or representation of a measurement. An *interpretive challenge* disputes an inference, model, validity domain, uncertainty report, or the equivalence asserted by a translation. An *institutional challenge* concerns authorization, exclusion, misuse, contributor rights, benefit distribution, or the adequacy of remedies. A single dispute may involve all three; procedures should allow linked but nonidentical assessments rather than forcing them into a generic “data quality” field.

An illustrative dispute record might link a contested object z to the claim being challenged, authorized challengers, submitted evidence, independent reviewers, current restriction on consequential use, interim notice, disposition, and available appeal. A decision to preserve an original record need not imply continued endorsement of its inference. A correction need not erase a history of earlier use where that history is relevant to responsibility. At the same time, indefinite public retention of personally identifiable or ecologically sensitive records cannot be defended by an unqualified appeal to provenance. Whether versioned information is retained, redacted, restricted, or erased depends on an applicable rights and preservation framework.

Proposition 8.1 (Technical traceability does not entail effective remedy). *Suppose a federation requires every permitted data transformation to be recorded in a complete, correctly linked provenance log. Without additional assumptions concerning challenge rights, review authority, and operational remedy, completeness of that log does not imply that an affected participant can obtain a correction, cease unauthorized reuse, or receive an effective remedy.*

Proof. Consider two otherwise identical data-processing systems with complete and accurate logs. In the first, an affected contributor can initiate a review, a competent body may amend an erroneous inference and suspend harmful reuse, and that action is communicated to dependent services. In the second, records are fully visible but no participant is empowered to challenge use or alter an endorsed output. The systems satisfy the same traceability premise but have different remedial capabilities. Consequently the premise alone is insufficient for the stated conclusion. $\square$

Proposition 8.1 is a logical non-implication, not an empirical finding about WMO, Argo, or another existing programme. It motivates a separate review of *remedy in operation*. A federation should test whether complaints can be brought, whether relevant evidence can be understood, whether independent and culturally appropriate review is available where required, whether corrective decisions propagate to dependent applications, and whether those affected can verify the practical result. Institutions may need plural review forums: scientific review for a model discrepancy, legal or community procedures for access restrictions, and operational incident procedures for a service failure. Overlapping jurisdictions create difficult coordination questions that a metadata standard cannot settle by itself.

Repair may restore a productive relationship, but restoration need not be equated with returning the system to its previous configuration. A correction can reveal a previously excluded contributor, alter the representation of an ecological process, or require changes in how authority is assigned. It may modify

the conditions for further cooperation. This is one reason to retain a relational and historically sensitive account of governance. Whether a particular institutional repair is legitimate remains a contestable normative question; the architecture should expose the relevant decision process rather than encode one theory of justice as a universal automated output.

8.5 Membership, Access, and Exit

A federation must decide who can publish, discover, compute, review, govern, and withdraw. These are distinct capacities. Membership in an intergovernmental observation network may be attached to national responsibilities, while an indigenous community, small laboratory, civil-society observer, or independent scientific consortium may contribute important observations without occupying the same formal institutional position. A common technical interface cannot itself determine who possesses legitimate authority over a sensitive dataset or who ought to have a voice in setting the rules of its reuse.

We therefore propose an explicit, role-specific participation map rather than a binary inside/outside classification. For institution or community i and activity a in governance state g , let

$$\text{Role}_g(i, a) \in \{\text{contribute, discover, access, compute, review, govern, restrict, exit}\}, \quad (39)$$

where this notation denotes categories of recognized relations and permissions, not an exhaustive or mutually exclusive list. An actor may be eligible to contribute data but unable to publish sensitive raw measurements; may review the scientific use of a record yet lack authority to disclose it; or may withdraw from one activity while retaining legal or ethical obligations relating to previous uses. These differences should be auditable at the institutional interface, with the substantive authority determined by context-specific agreements and applicable rights.

Exit is particularly important when data and models are copied across borders. There is no technical operation that can guarantee restoration of a pre-sharing state after unrestricted replication, publication, training, and downstream reuse. An honest design should distinguish termination of future provision, revocation of credentials, restrictions on future use, feasible deletion of controlled replicas, retention under lawful archival duties, and obligations concerning previously generated derivative products. Promises of complete revocation are inappropriate when their technical or legal feasibility has not been established. Data contributors should be informed of these limitations when deciding whether to participate.

A governance arrangement that permits exit only in principle may also reproduce power asymmetries. Less-resourced organizations may depend on one powerful platform for storage or scientific visibility; the expense of migration can make exit practically unavailable. A proposed mitigation is to specify portable metadata and export formats, continued access to provenance and audit records as permitted, alternative scientific service providers, and independent mechanisms for negotiating disputed restrictions. These arrangements involve costs, competing rights, and potential security risks. Their design requires evaluation rather than a presumption that portability is automatically beneficial in every setting.

8.6 Transnational Asymmetry and Shared Rule Formation

International observation infrastructure is supported by unequal distributions of sensors, scientific expertise, network links, computing facilities, energy, legal resources, and linguistic reach. Nominally equal publication rules can impose unequal real burdens. A participant that supplies frequent observations may lack the infrastructure to perform the resulting analyses, while another institution may acquire substantial computational and commercial advantage from data produced elsewhere. These differences matter even when access conditions are formally open and when every user can query the same registry.

It is therefore necessary to distinguish *rule interoperability* from *participatory rule formation*. Interoperability makes it possible to implement declared standards across sites. Participatory rule formation concerns who can propose, amend, challenge, and interpret those standards, who bears compliance costs, and whose evidence or knowledge is recognized when interpretations conflict. The first problem is partly

technical; the second involves authority, capacity, and political economy. Ostrom’s analysis of polycentric systems cautions against the presumption that one institutional model fits diverse settings, while UNESCO’s Recommendation on Open Science emphasizes inclusive cooperation, capacity-building, and differentiated legitimate restrictions [52, 57].

Table 12: Proposed governance responsibilities and possible asymmetries in a global scientific federation.

Layer	Candidate institutional practice	Asymmetry to assess
Observation	Shared calibration, maintenance support, contributor records	Unequal sensor coverage and repair capacity
Discovery and access	Transparent core conditions and justified restrictions	Differential bandwidth, licensing, language access
Computation	Portable workflows, resource sharing, reviewable usage terms	Concentrated GPU/energy and model-development capacity
Interpretation	Joint validation and plural scientific assessment	Unequal credibility and authority to define targets
Governance	Accessible rulemaking, incident review, independent challenge	Costs of representation and ineffective appeal
Ecological responsibility	Impact accounting, mitigation and remedial procedures	Externalized burdens and absence of affected communities

The proposed comparisons in Table 12 identify areas for institutional assessment rather than rankings of any country or programme. A global infrastructure could require shared minimum commitments without centralizing the substantive authority to decide every data-use dispute. Conversely, decentralization alone cannot prevent extraction, exclusion, or domination by actors who control indispensable computing, funding, or access routes. Public and commons-based arrangements require scrutiny of the same differences between nominal governance rights and actual capacity to exercise them.

An empirical research design should examine who participates in agenda-setting, who performs observation work, which contributors are credited, who develops the highest-value models, where environmental burdens arise, and how often challenges result in effective changes. It should also include failures and withdrawals, rather than selecting only successful instances of scientific cooperation. Without such evidence, a statement that a federation is globally cooperative may amount to an institutional aspiration rather than a demonstrated property.

8.7 Common Norms and Normative Interoperability

Different scientific and institutional participants can receive an identical observation while assigning it different meanings and practical implications. Their interpretive systems may have been formed through different relational histories: laboratory traditions, environmental stewardship, legal obligations, institutional missions, access to computing, and experience of previous disclosures. This is a possible origin of epistemic and normative conflict. It does not justify assuming that all disagreements are symmetrical, that all interpretations are equally well supported, or that only cross-system translation can resolve them. Some disputes can be settled through better evidence; others concern legitimate authority, conflicting rights, or incompatible purposes.

The search for some common workable norm is not a novelty of the present inquiry. Diplomatic practices described as *qitong cunyi* (求同存异, seeking common ground while preserving differences), Rawls’s overlapping consensus, legal-pluralist scholarship, boundary-object research, and theories of norm contestation all offer important predecessors [53, 48, 54, 60]. They also indicate limits: practical

agreement may be grounded in divergent reasons, can coexist with distinct normative orders, and can later be challenged. International-law sources provide another comparison, but not a replacement for local authority. Article 38 of the ICJ Statute identifies sources and subsidiary means for the judicial determination of international law, within its specific jurisdictional setting [40]; it does not imply that a technical federation may create universally binding law.

For the narrower purpose of a knowledge federation, we propose *normative interoperability* as an investigative label for the ability of differently situated participants to identify, formulate, apply, contest, and revise a limited cooperative rule even when their interpretations and justifications differ. This is distinct from epistemic interoperability: understanding an observation or the reasons behind another institution's judgment need not entail accepting its normative conclusions. Conversely, accepting a shared disclosure protocol need not establish identical interpretations of the ecological evidence. The protocol's scope, applicability, and legal effect must be separately established.

Let P be a specified *nonempty* set of participating institutions with an explicitly documented basis for their role in the proposed protocol, e a shared or partially accessible evidence dossier, h a recorded interinstitutional history, and $\mathcal{N}(e, h)$ the candidate protocols considered in a declared decision context. Let $a_i(n \mid e, h, \iota_i) \in \{0, 1, ?\}$ record institution i 's stated acceptance of rule n , where $?$ denotes unavailable or unresolved acceptance rather than consent. A provisional working set is

$$\mathcal{W}_P(e, h) = \{n \in \mathcal{N}(e, h) : a_i(n \mid e, h, \iota_i) = 1 \text{ for each } i \in P\}. \quad (40)$$

The membership of $\mathcal{N}$, the recognition of P , and the meaning of assent are historical and institutional choices. Equation (40) is a descriptive candidate-screening device, not a universal social-choice rule and not a test of legal validity or justice. The candidate space can be expanded by negotiation, experimentation, translations, or newly admitted participants. It need not equal the static intersection of inherited legal or moral systems. Membership in $\mathcal{W}_P$ asserts only that the *recorded* participants passed this acceptance screen for the selected protocol version; it is silent about omitted or unrecognized affected persons, coercion, required procedural safeguards, jurisdiction, legal effect, and ethical justification. A later revision to P , to the meaning or validity of an acceptance record, or to the protocol version requires a new assessment rather than inheritance of an earlier membership result.

Proposition 8.2 (Interpretive difference is compatible with bounded practical agreement). *Non-equivalence of interpretations of a shared record does not, by itself, imply that the working set in Equation (40) is empty. Nor does the mere existence of shared evidence imply that it is nonempty.*

Proof. For the first assertion, stipulate two participants interpreting a coastal observation differently: one prioritizes scientific reproducibility, the other protection of a sensitive habitat. Both nevertheless accept an authorized research channel with restricted public coordinates, so at least one candidate belongs to their working set. For the second, retain the shared record but stipulate that the participants reject every candidate protocol in the declared set because their requirements are incompatible. Both constructions are possible without changing the fact of access to shared evidence. The claim concerns logical possibility, not the frequency or normative desirability of either outcome. $\square$

These basic logical distinctions establish no general claim of originality over agreement amid difference. The proposed research contribution lies in specifying and testing an *interface* between interpretive divergence and institutionally implemented cooperation: an interpretation-and-evidence dossier; statements of distinct reasons and reservations; the decision-maker's competence and applicable legal bases; an inspectable protocol version; participation and review rights; observable performance and failure; and mechanisms for amendment, withdrawal, and remedy. Such records could support empirical investigation into whether norm changes alter the capacity to recognize or challenge claims, and how performance histories change subsequent practical reliance.

A further limit is necessary. Agreement among admitted participants is insufficient evidence of legitimacy when affected communities, future users, or ecological interests were excluded; coerced assent, unequal exit capacities, and incompatible legal duties can also defeat a simplistic acceptance test. The

scientific justification of a measurement, the positive legal validity of a rule, participant acceptance, and the ethical legitimacy of outcomes remain separate assessments. Proposed cross-system nomos must therefore remain open to contestation and, where justified, transformation. The companion study can examine how repeated rule performance, conflict and repair influence historically situated trust. For this paper, the result is a bounded governance-research question and a candidate architecture interface, not a solved theory of transnational norms.

The common-language problem is deeper than a choice of vocabulary. The participating GR systems may have formed distinct generalized languages through their own relational histories, including different conventions for evidence, personhood, authority, obligation, harm, and adequate remedy. A limited common norm may become manifest in shared protocol text and simultaneously acquire relational efficacy through actual reliance, compliance, criticism, and repair. These are analytically distinct qualifications: the existence of a signed agreement is a manifestational fact, whereas its stable efficacy within relations is an additional claim requiring evidence; the agreement's ultimate moral validity is not inferred from either fact. This follows the coupled reality-register account in Section 2.4. Normative interoperability should accordingly record which language-bridges are task-specific, which differences remain, and whether affected participants can change the agreed forms. Continued convergence is neither assumed nor always desirable.

The engineering profile of Equation (18) is useful here only as an *accounting interface for attributed assessments*. In particular, a scientist's acceptance of an inferential mapping, a participant's assent to a proposed cooperative protocol, a competent institution's authorization of a specified operation, and an affected party's capacity to contest that operation are four distinct matters. None can be silently inherited from another when records cross jurisdictions or interpretive contexts. Where a claimed authority is disputed, the system should record the competing grounds and the relevant review forum instead of fabricating one effective global decision. The fuller knowledge-circulation contract is developed in Section 10.10.

8.8 Boundaries to Companion Paper V on Generative Trust

The present section supplies a *governance specification* for global observation and knowledge infrastructures. It identifies relations among participants, verifiable records, commitments, access permissions, conflicts, and remedies. It does not establish a general metaphysical or sociological theory of trust, derive one universal measure of credible cooperation, or prove that any proposed federation would increase trust. Its formal expressions are limited to typed research objects and logical distinctions, rather than a quantitative trust estimator or optimized institutional constitution.

The proposed companion paper on generative trust and transnational governance can investigate problems that are substantially wider than this architecture: how trust relations emerge through recurrent interaction, how asymmetries alter the meaning of reciprocity, how institutional failures change possibilities for further cooperation, how remedies generate or suppress new capacities for participation, and how multiple legitimate authorities can coexist without eliminating disagreement. It should be open to analyses of power, extraction, epistemic bias, conflict, and the conditions under which trust is misplaced. This independence prevents the present study from assuming that trustworthy relations are merely an output of infrastructure design.

For this paper, the practical research requirements are narrower and testable: each consequential use of shared knowledge should make relevant evidence, scope, rights, uncertainty, and authorization inspectable by legitimate reviewers; affected participants should have access to appropriately designed correction and remedy procedures; and commitments about access, reuse, acknowledgement, and dispute resolution should be evaluated through their actual operation. These conditions do not determine a single globally optimal institutional order. They provide concrete questions for comparative investigation across meteorological, oceanographic, ecological, and other scientific networks.

Finally, the relation to the first paper of the programme should remain carefully bounded. Improved trust and credible coordination might reduce particular monetary coordination costs, but may also sup-

port the creation of additional credit, contracts, and costly obligations. A global knowledge infrastructure can transform the conditions for economic exchange without thereby proving distributed monetary dissolution [37]. The sequence of research is accordingly methodological: the present paper studies the conditions for shared observation and inquiry; later studies will examine ecological exploration, productive infrastructures, and the generativity of trust in their own right. Section 9 will next consider public knowledge, community rights, and ecological obligations in a more explicit normative and environmental register.

9 Public Knowledge, Rights, and Ecological Obligations

The preceding discussion described federated observation, translation, assimilation, and relations of credible cooperation. The present section addresses a further question: what makes a shared knowledge infrastructure *public* in its purposes and consequences, while its constituent observations, productive capacities, and affected relations remain heterogeneous? Public purpose cannot be read directly from the number of downloadable records, the size of a repository, or the availability of an AI interface. Infrastructures can expand scientific opportunity while reallocating interpretation, extraction, environmental burdens, and rights of decision to actors far removed from the sites where knowledge is produced. The objective here is to establish a bounded vocabulary for these distributions and their associated obligations. A comprehensive theory of ecological value, Indigenous law, or transnational justice lies outside the present architecture paper; the discussion identifies requirements that later research must examine rather than purporting to decide them universally.

The distinction is particularly consequential for environmental observations. Weather records may be used for warnings, agriculture, or research; marine acoustic recordings may contain scientifically important signals along with commercially sensitive operational information; precise records of threatened species may support conservation analysis but also expose vulnerable populations to disturbance or collection. Even the reproduction of an otherwise unrestricted dataset may depend on electricity, water, physical equipment, and labour whose costs are not captured in an access licence. Thus the proposed federation must represent several non-equivalent questions: *who may access which records, for which tasks, on what authority, at whose expense, and through which procedures for review and repair?*

9.1 Public Purposes and Differentiated Data Access

A public knowledge infrastructure should make its purposes inspectable, not simply declare all data to be public. Scientific reproducibility, environmental monitoring, rapid disaster warning, education, local self-determination, and accountable public decision-making are partially overlapping purposes, and they may motivate different access requirements. For instance, an emergency forecast may require broad and timely dissemination of a warning derived from a restricted source without requiring unrestricted publication of the source's fine-grained observations. Likewise, a reproducibility request may be satisfied through independent audit, secure analysis, or a carefully constructed aggregate where unrestricted source-data release would cause avoidable harm.

The FAIR principles concern findability, accessibility, interoperability, and reusability, including machine-actionable metadata and specified access protocols; they do not imply that every dataset must be released without conditions [61]. UNESCO's 2021 Recommendation on Open Science encourages access that is as open as possible, subject to proportionate and justified restrictions including privacy, relevant knowledge rights, confidentiality, and protection of endangered species [57]. These sources motivate a differentiated-access design, but neither resolves every legal or ethical conflict in an international federation. A locally authorized custodian may have obligations that do not reduce to a globally uniform rule, while common metadata and review procedures can still make those obligations intelligible across nodes.

For an observation product d , a requesting actor a , a declared scientific or institutional purpose p ,

and a situated governance context g , represent an access decision as

$$\text{Access}(a, d, p \mid g, h) \in \{\text{public, mediated, restricted, denied, pending}\}, \quad (41)$$

where h denotes the relevant authorization and revision history. Equation (41) describes potential decision outcomes, not a globally binding permission algorithm. In particular, *mediated* access might involve an institutional data enclave, on-site inspection, a supervised computational query, or a purpose-limited research agreement. The presence of an accessible data catalogue must not be confused with permission to acquire, re-identify, combine, or republish the underlying records. Nor does *denied* imply that the holder may refuse an otherwise legally required disclosure. Actual powers and duties depend on the jurisdiction, source, and institutional mandate.

A use-relevant authorization record may contain $(a, d, p, \ell, c, \sigma, r, h)$, with an access level ℓ , conditions c , authorizing body or process σ , review channel r , and history h . If a governing institution cannot explain the basis or boundaries of a decision, interoperability alone will not make it legitimate. Conversely, a justified restriction should not automatically suppress public metadata about the existence, provenance category, scientific validity, permissible uses, and correction status of a dataset. The particular metadata fields that can be disclosed may themselves require protection in sensitive cases. The architectural question is therefore how to preserve as much useful discovery and independent scrutiny as the applicable rights and risks permit.

Table 13: Candidate access arrangements for heterogeneous public-knowledge purposes; the rows describe research options, not a universally applicable legal classification.

Example record or task	Candidate access arrangement	Necessary additional examination
Wide-area meteorological warning	Public products and relevant core observations	Forecast reliability, update obligations, timely public communication
Community-associated ecological record	Purpose-limited sharing, community-governed review	Authority, collective benefit, consent, acknowledgement, remedy
Sensitive species location	Generalized public record; protected precise location	Species-specific threat, analytical distortion, escalation and review
Restricted environmental microdata	Secure analysis, approved aggregates, mediated audit	Re-identification, derivative outputs, responsible reviewers
High-cost coupled model ensemble	Shared workflow and negotiated computation	Electricity, water, access capacity, cost allocation, reproducibility

Table 13 makes a basic point: openness may be a property of a *publicly usable scientific result*, a reproducible procedure, or a governance record, even where direct distribution of every contributing input is inappropriate. The applicable balance should remain challengeable as scientific methods, threats, or legitimate community decisions change. Proportionate access is thus an ongoing, relational arrangement, not a once-for-all label in a static repository.

9.2 FAIR, CARE, and Community Authority

The FAIR and CARE frameworks address partly different dimensions of research-data governance. FAIR specifies features that improve technical discovery, exchange, and reuse. The CARE Principles for Indigenous Data Governance were articulated through Indigenous data-sovereignty scholarship and networks; their components are *Collective Benefit, Authority to Control, Responsibility, and Ethics* [6]. This genealogy must not be diluted into a generic slogan of “ethical data.” Indigenous Peoples’ rights and

authority over data about their communities, cultures, and territories are substantive institutional issues; they are not satisfied merely by adding attribution fields to a record.

In a proposed international federation, the two approaches can inform separate but connected questions. A record may be technically findable and reusable yet subject to authority or collective-benefit obligations that an external repository fails to recognize. Conversely, respecting community authority does not preclude developing high-quality metadata, interoperable records, and locally chosen sharing protocols. The conflict to investigate is not simply *openness versus closure*, but whether a particular infrastructure allows affected communities to determine or meaningfully contest the purposes, conditions, interpretation, and benefit distribution of uses involving their data.

The federation should therefore record at least four distinctions. *Provenance* concerns the documented origins and transformations of data; *contribution recognition* concerns the people and institutions that helped generate the knowledge; *authority* concerns legitimate decision-making powers over use and disclosure; and *benefit or remedy* concerns the effects of circulation and the ability to respond to harm. No one item entails the others. Crediting an Indigenous observer does not itself constitute consent to a later commercial model-training use. A formally permissive licence may coexist with historical patterns of extraction or with collective interests that require an independent legal and ethical assessment.

Consider a hypothetical marine observation project whose datasets combine instrument readings with locally held ecological knowledge. The instruments might be administered by a university, the knowledge generated through the long-term activities of an Indigenous community, and the data subsequently used in a third party's forecasting model. Source recognition, public-interest justification, contractual rights, and community authority may point in different directions. A responsible governance design would first establish who actually holds the relevant rights under the applicable arrangements, how participation and authorization have been documented, and what procedures exist for challenging inferred classifications or unintended secondary uses. This example does not assume one uniform legal regime for all community data or treat community groups as internally homogeneous.

A candidate governance dossier can be written as the ordered tuple

$$\mathcal{D}(d, p, h) = (\mathcal{P}, \mathcal{A}, \mathcal{C}, \mathcal{B}, \mathcal{Q}; d, p, h), \quad (42)$$

where $\mathcal{P}$ denotes provenance and contributor recognition, $\mathcal{A}$ recognized authority, $\mathcal{C}$ use conditions and consent where applicable, $\mathcal{B}$ claimed or observed distributions of benefit and burden, and $\mathcal{Q}$ processes for questions, revisions, and remedy. The tuple in Equation (42) is a *descriptive information object*, not a claim that these heterogeneous responsibilities can be collapsed into a welfare score or machine-verifiable compliance certificate. It should also carry an explicit indication when authority, consent, or evidence remains disputed or unknown.

An infrastructure that records such dossiers can assist good practice but cannot manufacture authority where none has been recognized, identify rightful representatives without legitimate processes, or infer collective consent from a statistical majority. These problems require institutional participation and context-specific inquiry. The more general challenge for public knowledge infrastructure is to provide forms of cooperation that preserve meaningful community governance while enabling scientific findings to be evaluated and, where appropriate, shared beyond the institution that first records them.

9.3 Scientific Capacity, Benefits, and Interpretive Agency

Access to downloadable data and access to the *capacity to make consequential interpretations* are different. Scientific benefit depends partly on instruments, reliable connectivity, computing, skilled labour, funding, language resources, model documentation, and the authority to question resulting analyses. A participant may collect valuable measurements yet lack the computing power to reanalyse the assimilated product derived from them. Another participant may control the registries, reference ontologies, model-service endpoints, and publication channels that determine which interpretations become visible. In such a system, apparent openness can coexist with substantial inequalities in scientific agency.

For an institution i and a specified inquiry q , let

$$\mathcal{K}_i(q) = (O_i, X_i, M_i, I_i, V_i, R_i)_q \quad (43)$$

represent a qualitative or separately measured profile of observation capacity, usable access, modeling and computational capacity, interpretive participation, independent validation, and effective routes for revision or remedy. Equation (43) does not define a scalar index: different dimensions may be non-commensurable, constrained by distinct institutional rights, or subject to trade-offs. It permits a more informative comparison than counting whether each institution possesses an account on the same platform.

Benefits should likewise be investigated as heterogeneous effects rather than assumed to follow automatically from data reuse. For a particular cooperative activity, plausible outcomes include scientific understanding, improved warning capacity, the ability to train and retain local researchers, financial income, recognition of observation work, ecological protection, and the capability to contest external decisions. Costs can include staff time, equipment upkeep, loss of control over local interpretations, computational expenditure, and environmental harm. A comparison should identify who receives each benefit, who bears each cost, and what counterfactual alternative is being assessed; it should not silently treat an aggregate gain in available information as proof of distributive legitimacy.

The UNESCO Open Science Recommendation calls for inclusivity, capacity building, and cooperation while recognizing legitimate restrictions [57]. Such norms can guide an inquiry into participatory capability, but actual progress must be observed. Examples of testable indicators include the extent to which institutions contributing measurements can reproduce analyses, participate in model design, obtain documented attribution, request revisions, access computational support, and successfully challenge an erroneous classification. Programme membership and the number of downloads are insufficient substitutes. These indicators must be interpreted in the organizational circumstances where they are recorded, including whether participants can exercise formal rights without prohibitive expense or risk.

Cooperation can also involve capacities that remain outside a single organization's possession. Researchers may form networks that share instrument repair, high-performance computation, and independent calibration through reciprocal arrangements. The objective should be to examine how these relations change the feasible space of collective inquiry, not to prescribe self-sufficiency for every node. There may be substantial value in interdependence when it is transparent, contestable, and supported by credible commitments. But concentration of indispensable services can turn that interdependence into a source of asymmetric control. The architecture therefore needs a record not only of accessible outputs but of the critical dependencies through which knowledge is made and interpreted.

9.4 Environmental Sensitivity and Conservation Risks

Ecological knowledge is often most valuable when observations are precise, but precision can sometimes increase the risk of damaging the object of inquiry. Sensitive biodiversity records are a concrete example. GBIF guidance recommends considering generalization or withholding of fine-scale occurrences where publication could expose a species to exploitation or disturbance, and documenting the transformation and associated uncertainty [8, 26]. The objective is not to conceal all ecological observations but to distinguish legitimate scientific use, appropriately limited disclosure, and unnecessary exposure. Sensitivity is contextual: the same coordinate precision may be innocuous for a common species in one region and harmful for a threatened population in another.

One response is to maintain protected source observations with role- and purpose-dependent scientific products. Let s be a sensitive observation and T_g a documented generalization under conditions g . The public record may be $\tilde{s} = T_g(s)$, accompanied by transformation metadata and an account of the loss of spatial resolution. The analytic question is whether an intended target quantity ψ can be estimated adequately from $\tilde{s}$. For some targets, coarse occurrence data may remain useful; for others, necessary fine-scale structure is discarded. This repeats the noninvertibility concern of Section 6, now with conservation and disclosure reasons for retaining a deliberately lossy transformation.

Proposition 9.1 (Protection by suppression does not ensure harmless derived inference). *Suppose a protected observation s is withheld, but a publicly released derived output $y = F(s, z)$ and public auxiliary data z together permit an authorized or unauthorized observer to recover a protected attribute $\phi(s)$. Then withholding s alone does not protect $\phi(s)$ against inference from (y, z) .*

Proof. By hypothesis, a reconstruction procedure R satisfies $R(y, z) = \phi(s)$ for the accessible output y and auxiliary information z . The observer can therefore obtain $\phi(s)$ without direct access to s . The conclusion concerns this stipulated reconstruction and is not a general assertion that all outputs are invertible. $\square$

Proposition 9.1 is elementary but relevant to AI-assisted inference. A model trained on protected ecological positions may reproduce or sharply localize them through predictions, explanations, generated maps, or repeated queries even if the original records remain behind access controls. Whether leakage occurs depends on the scientific target, model, threat actor, auxiliary information, and disclosure mechanism. The federation should consequently examine *derivative disclosure* and model-query risks, rather than assume that restricting raw-data downloads ends the problem. The same caution can apply to sensitive human or community-associated environmental records, subject to their distinct rights and risks.

Environmental sensitivity also extends beyond secrecy. Poorly designed monitoring campaigns may disturb habitats, generate underwater noise, introduce equipment or other materials, or divert local scientific labour. The evidence required to evaluate such harms differs by project, ecosystem, and technology. A global observation network may support conservation only when the ecological effects of collecting, processing, using, and acting on observations are themselves assessed. Later work on marine robotics will investigate these empirical questions independently, with particular attention to scientific discovery as a purpose in its own right.

9.5 AI Interpretive Concentration and Infrastructural Power

AI systems may make heterogeneous observations more searchable, perform rapid quality checks, suggest semantic alignments, synthesize large evidence collections, or assist modeling across languages and disciplines. Such capabilities could expand participation, yet their organization may also concentrate interpretive authority. If a small number of model operators controls embeddings, retrieval rankings, model updates, evaluation datasets, or access to large computational facilities, users may increasingly encounter a privately or institutionally filtered representation of global scientific observations. This outcome is possible even when the underlying records remain publicly available in principle.

The research problem is a distribution of *epistemic affordances and governance powers*, not a presumed intention of AI developers. A model may omit minority interpretations because of sparse training data, incompatible representations, optimization objectives, or inaccessible source collections. The loss can be amplified when other institutions consume its summaries without examining the originals. Cross-system epistemic translation, as developed in Section 6, is therefore not merely an accuracy problem; the ability to challenge, compare, and revise translations also affects who can contribute to knowledge formation.

An audit should distinguish source-level access, model-level access, influence over training or configuration, computational affordability, control of output distribution, and the possibility of independent validation. A federation can expose model provenance, training-data declarations where legally and technically possible, version histories, uncertainty limits, and routes to contest a result. Yet such documents alone cannot guarantee meaningful review. A challenge procedure may be nominally available while the affected observer lacks the language resources, expertise, money, or standing needed to use it. Review must therefore be evaluated in operation.

The following conceptual separation can guide empirical tests:

$$\mathcal{I}_i(q) = (S_i(q), C_i(q), E_i(q), A_i(q)), \quad (44)$$

where S_i denotes available source evidence, C_i effective analytical capability, E_i influence over interpretive and evaluation procedures, and A_i opportunities to appeal, repair, or independently verify outcomes. Equation (44) again records dimensions rather than a universal power index. The distinction permits evaluation of situations in which open scientific data and unequal authority over consequential interpretations coexist.

We should also resist an inverse simplification: federated or open-source implementation does not automatically produce epistemic justice. Biases may be distributed across many nodes, and common standards may entrench the categories of historically dominant institutions. Similarly, centralized services may sometimes improve accessibility or emergency responsiveness. What matters for this investigation is the empirically and institutionally grounded configuration of scientific capabilities, rights, dependencies, participation, correction, and alternatives. The desirability of any specific arrangement cannot be inferred from decentralization alone.

9.6 Data, Computation, and Energy Costs

Global observation and knowledge infrastructures also depend on material production. Instrument manufacture, satellite operation, ship expeditions, sensors, network transmission, storage, model training, reanalysis, inference, cooling, maintenance, and equipment replacement can require substantial energy and materials. A proposed federation cannot be evaluated as environmentally beneficial simply because its application domain is environmental observation. More complete scientific knowledge may enable valuable conservation or adaptation, but the infrastructure’s own footprint and its induced uses require separate analysis and appropriate counterfactuals.

For an explicitly bounded infrastructure $\mathcal{N}$ and accounting period or event set h , define a noncommensurate material and ecological account

$$\mathcal{E}_{\mathcal{N}}(h) = (E_{\text{electric}}, W_{\text{water}}, M_{\text{materials}}, L_{\text{land}}, Z_{\text{ecological}}; \partial\mathcal{N}, h). \quad (45)$$

Here $\partial\mathcal{N}$ specifies the chosen accounting boundary, and $Z_{\text{ecological}}$ denotes a context-specific dossier of ecological effects, evidence, uncertainty, and affected receptors rather than a single exchangeable damage number. Equation (45) does not purport to be a life-cycle assessment standard. At minimum, comparisons should clarify whether values refer only to data-center operation, to the wider electricity system, to embodied equipment and construction, or to an entire observation-to-analysis pipeline. The same numerical energy figure can have substantially different interpretation under different boundaries and electricity mixes.

Established policy and measurement work supplies relevant but limited starting points. The European Union’s Delegated Regulation 2024/1364 specifies, among other indicators, power usage effectiveness and water usage effectiveness for reporting within its defined scope [17]. The International Energy Agency’s updated 2026 analysis estimates data-center electricity use of approximately 485 TWh in 2025 and projects about 950 TWh for 2030 under its central outlook; these quantities concern the sector, not the footprint of the federation imagined here [41]. They contextualize the scale of the wider computing environment but do not authorize attribution of a particular share of consumption to any proposed global knowledge infrastructure.

A responsible decision requires at least a counterfactual account of benefits and environmental consequences, not an assumption that minimizing electricity per computational task maximizes ecological performance. A lower-energy model might require additional sensing in a fragile habitat; a more energy-intensive reanalysis might improve early warning and reduce other harms. Conversely, a highly efficient data center can enable large increases in computation or shift environmental burdens elsewhere. These relationships must be measured and analyzed with explicit boundaries, uncertainty, and alternative scenarios. Table 14 separates the required domains of evidence, rather than forcing heterogeneous environmental and epistemic consequences into an unexamined common currency.

Finally, ecological obligations need institutional representation. Ecosystems and nonhuman organisms do not ordinarily participate as conventional contracting parties in data-governance arrangements,

Table 14: Ecological assessment boundaries and evidence needs for proposed knowledge infrastructures.

Assessment domain	Evidence to collect	Limit or counterfactual to specify
Electricity and cooling	Metered energy, local grid conditions, equipment utilization	Operational versus full-system accounting; alternative compute allocation
Water and materials	Water withdrawal/consumption, equipment lifespan, procurement	Local water stress, embodied costs, replacement schedules
Ecological disturbance	Site baseline, habitat effects, species exposure, monitoring	Seasonal variation, cumulative effects, alternatives to deployment
Knowledge benefit	Warning skill, discovery, valid reuse, local capacities	Additional value compared with existing observation and models
Governance and remedy	Affected actors, review decisions, mitigation, documented repair	Whose objections count, effectiveness of remedial action

yet they can be affected by observation activities, AI infrastructure, and actions enabled by shared knowledge. The institutional task includes scientifically evaluating effects, identifying responsible decision processes, incorporating appropriate representation and stewardship, and making mitigation and remedies reviewable. This is not resolved by declaring an ecological observation dataset open or by balancing financial accounts. The earlier monetary-dissolution paper similarly distinguishes financial clearing from ecological incidence; the connection here remains conceptual and does not constitute empirical confirmation of either research programme [37].

Section 10 will specify the common-knowledge and generalized-translation infrastructures needed to make these concerns operational. Section 11 will then test the resulting claims in explicitly hypothetical federation scenarios. In particular, it will ask whether a design can support a valuable public scientific result while restricting harmful disclosure, recognizing affected contributors, recording environmental burdens, and retaining workable procedures for contestation and revision. Such tests are needed before any claim that a global knowledge infrastructure is simultaneously interoperable, publicly beneficial, and ecologically responsible.

10 Common Knowledge Creation, Circulation, and Generalized Translation Infrastructures

The preceding analysis established why global observation, publicly useful knowledge, rights, and ecological responsibilities cannot be collapsed into a problem of exchanging data. The objective of the present section is more constructive: to identify *what would have to be built* for heterogeneous participants to create knowledge together, circulate its products and productive capacities, and mediate the passage of representations, practices, obligations, or heterogeneous values across distinct generative-relational (GR) contexts. An internationally accessible repository is a necessary component for some purposes, but it cannot independently carry out these activities. A broader infrastructure requires instruments, computational resources, research environments, translation interfaces, legal and institutional procedures, maintenance, education, and mechanisms through which excluded or affected parties can intervene.

This discussion is a *candidate infrastructure specification*, rather than a feasibility finding or an assertion of a universal ontology. Existing standards and institutions offer selected building blocks, while the proposed synthesis, especially generalized relational and value mediation, remains unimplemented and demands a separate engineering and institutional investigation. The ocean provides an illustrative

experimental domain; the architecture is not restricted to marine research. The present section articulates interface requirements and failure tests. A companion investigation could develop them into specifications, implementations, and comparative trials without prematurely treating their performance as established.

10.1 Knowledge Products, Generative Capacity, and Circulation

Three infrastructural objects should be distinguished. A *knowledge product* is a bounded record or artifact such as an observation, dataset, model, argument, method description, publication, or software release. A *knowledge-producing activity* includes situated measurement, questioning, inference, collaborative experimentation, criticism, teaching, or institutional review. A *generative capacity* is a task-relative ability to undertake further such activities, sustained by skills, instruments, methods, legal permissions, accessible materials, computation, energy, and relationships with other participants. These are analytical distinctions: one artifact may also function as a tool, and capacities may transform the activities and categories through which they are described. Their separation prevents formal availability of records from being mistaken for effective public scientific participation.

A common production infrastructure should therefore connect at least six services. First, *question and experiment workspaces* would preserve alternative hypotheses, research purposes, experimental designs, negative results, and routes for independent challenge. Second, *shared but differentiated research instruments* would allow observation requests, calibration, scheduling, maintenance, and safe operation to be coordinated without presupposing that every laboratory must share hardware. Third, *versioned scientific workbenches* would link source records, analytic code, model configurations, computing environments, and interpretable output histories. Fourth, *distributed discovery, publication, and circulation services* would support identifiable artifacts and provenance-preserving reuse, including justified restrictions. Fifth, *epistemic-capacity services* would enable access to teaching, repair, methodological assistance, computational allocation, and replication facilities. Sixth, *public critique and revision interfaces* would connect corrections and objections to the artifacts and activities they affect.

These functions can be composed without imposing uniform ownership, a universal curriculum, or mandatory openness. The FAIR principles support data stewardship and machine-actionable reuse; UNESCO's Recommendation on Open Science provides a broader public framework including infrastructure and participation. Their contribution to this proposal is normative and design-oriented, not proof that a world-scale, uniformly beneficial commons is feasible [61, 57]. A research system might allow a community-governed observational record to remain locally controlled while publishing a challengeable derivative, a permissible query interface, and provenance indicating the conditions of its production. What circulates may be a result, method, reproducible capability, supervised research opportunity, or negotiated service, with different rights for each.

For a participant i and research task q under a documented configuration h , distinguish *artifact accessibility* from *effective generative participation*:

$$\text{Readable}_h(i, d) \not\Rightarrow \text{CanGenerate}_h(i, q \mid d). \quad (46)$$

A participant might be able to download an ecological model yet lack the scientific training, instruments, licensed dependencies, affordable computation, or authority needed to interrogate and extend it. Equation (46) records a non-implication, not a universal binary definition of competence. The effective-participation condition of Equation (59) specifies further evaluation dimensions. The resulting political-economic question concerns who controls the means of producing, validating, circulating, and revising knowledge, including the power to determine what counts as an admissible contribution. This inquiry should encompass public, commons-based, commercial, and mixed production arrangements rather than predetermining one correct institutional form [51, 52].

10.2 Operational Layers of a Knowledge-Producing Commons

The design object includes both *circulation interfaces* and *production interfaces*. The first permit a declared resource, interpretation, or result to be located and used under specified conditions; the second permit an actor to formulate a question, obtain appropriate instruments or computation, execute and criticize a method, and publish a revision. They must be connected, since an otherwise reusable dataset may arrive without the means to reproduce the claim for which it is used. A commons in this operational sense is an arrangement of differentiated rights and capabilities, not an assertion that all content is unconditionally released or collectively owned. The institutions of production and circulation may be public, commercial, community-governed, or jointly administered, with their relative efficacy remaining subject to inquiry.

An initial implementation can distinguish six interoperating layers. The *resource-description layer* exposes a versioned catalogue of observations, software, methods, instruments and restricted discovery descriptors. The *research-execution layer* publishes reproducibility envelopes, including computational dependencies, environmental assumptions, inputs and known nonreproducible elements. The *capacity-access layer* accepts bounded requests for instruments, expertise, computing, calibration and training; it records actual provision separately from an entitlement or application. The *circulation layer* records copies, derivations, disclosures, supervised access and provenance-bearing publications. The *mediation layer* presents task-relative interpretation and translation claims for scrutiny. The *governance-and-response layer* routes authorization questions, disputes, withdrawal notices, corrections and claims of harm to accountable processes. These layers specify responsibilities and interfaces, not a centralized stack or a priority ordering among systems.

Table 15: Selected operational interfaces for knowledge production and circulation. The identifiers, decisions and outcomes remain institutionally contextual; the table is a research design rather than a new interoperable standard.

Interface	Required record or response	Discriminating test
Register an inquiry	Question, methodological alternatives, relevant evidential supports, responsible submitter	An unconventional but admissible question remains discoverable
Request productive capacity	Instrument or computation request, constraints, authorized allocator, actual allocation and barriers	A nominally eligible participant can execute an independent bounded task
Execute and review	Versioned workflow, input identities, uncertainty, reviewer objections and negative results	An adverse result remains interpretable and independently challengeable
Circulate a derivative	Source references, declared transformation, conditions of use and custody	Reuse is distinguishable from independent observation or approved transfer
Correct or withdraw	Changed record, reason, affected dependencies, known notification recipients and unresolved copies	A correction reaches identifiable dependants without claiming universal recall
Contest a claim	Objector's scope, disputed proposition, competent forum, response and remaining remedy	A challenge can alter a consequential record or document an unresolved refusal

A reproducibility envelope is evidence of how an experiment was represented and executed under declared supports. It is not a guarantee that another participant can acquire equivalent hardware, legally use protected inputs, or reproduce a phenomenon outside the stated boundary. Similarly, an instrument reservation describes planned access, whereas a service log can distinguish reservation, attempted use, completed measurement, and scientifically usable output. This distinction is necessary for evaluating infrastructure policies that formally promise equal access while allocating scientific capacity unequally. Records must also provide safe ways to protect confidential, personal, sensitive ecological or collectively governed information, including restricted descriptions, purpose-limited queries, and independently su-

pervised inspection where suitable.

An economic and institutional assessment must include operation and maintenance. Instruments require calibration and replacement; archives require stable custodianship; services require recovery procedures and security updates; methods require education and interpretive support. A viable federation should therefore document the parties responsible for these continuing capacities, changes of operators, funding or service dependencies, and exit or succession arrangements. The ability to export a dataset is weaker than the ability to continue a scientific practice after loss of a dominant provider. A graded, intermittent or offline participation pathway may be more inclusive than requiring every node to provide uninterrupted connectivity.

10.3 Generalized Translation between Heterogeneous GR Contexts

The translation infrastructure addresses a different object from an ordinary vocabulary converter. Across heterogeneous GR contexts, what requires mediation may be a measurement category, model, inferred rule, practice, scientific claim, institutional position, socially recognized relation, obligation, or value. A single state or country cannot be treated as internally homogeneous: research laboratories, administrative bodies, communities, firms, and individual participants within it can occupy divergent relational, epistemic, and normative contexts. Likewise, references to distinct relational spacetime configurations are theoretical descriptions of situated histories and relations, not claims that national borders partition physical spacetime into separate universes.

For systems i and j , a candidate task-bounded mediation record is

$$\mathfrak{T}_{ij}^{(q)} = (D_i, D_j, R_{ij}^{(q)}, \Phi_q, \Lambda_q, \mathcal{A}_q, \mathcal{H}_q). \quad (47)$$

Here D_i and D_j are declared *manifestational domains* to be related, rather than exhaustive representations of participating ontologies; $R_{ij}^{(q)} \subseteq D_i \times D_j$ is a potentially partial, many-to-many, contested correspondence; Φ_q contains task-specific tests and proposed preservation criteria; Λ_q records uncertainties, translation loss, excluded cases, and recognized non-equivalences; $\mathcal{A}_q$ records relevant authority, permission, and obligations without itself establishing their validity; and $\mathcal{H}_q$ records the provenance and revision dependencies of the attempted mediation. The tuple is an *analytical design profile*, not a claim that all participating concepts or values possess enumerable encodings. It complements the generalized-language profile in Equation (21) and the local preservation condition in Equation (22).

A mediation may *compare* representations, *transport* authorized artifacts, *reinterpret* evidence in a receiving practice, *negotiate* a shared operational protocol, or *co-create* previously unavailable distinctions and methods. These operations have different success conditions. Agreement about an observation does not ensure agreement about its explanation; agreement about an explanation does not confer authority to enact a policy; legal compatibility does not automatically establish ethical legitimacy. A partial bridge may deliberately preserve incompatibility, dissent, or refusal rather than supply an apparently fluent but unwarranted equivalent. Historically formed scientific boundary objects and trading practices offer relevant precedents for coordinating work without complete identity of understanding, but do not themselves specify a universal translation operator [54, 24].

The difficult question is what an AI-based discovery system can infer about another participant's generative representation from available manifestations, experimental interactions, and declared constraints. Candidate methods include structured model comparison, interactive elicitation, causal and counterfactual testing, program induction, active learning, and task-relative correspondence search. They can support hypotheses concerning another system's distinctions and procedures; they cannot guarantee unique recovery of hidden rules from finite observations. The noninvertibility counterexamples of Section 6 therefore remain foundational. A competent mediator must record underdetermination and avoid representing successful prediction of selected outputs as complete understanding of a person, community, legal order, or biological system [35].

10.4 Preservation Claims, Translation Failure, and Institutional Refusal

A practical translation service should publish *claims of correspondence* rather than silently assert semantic identity. For a bounded task q , let E_q be its declared test cases, p_q the property whose preservation matters, and r a candidate relation connecting expressions or activities in contexts i and j . A reviewer may ask whether paired cases satisfy a task-specific preservation requirement, such as equivalent observational units within calibrated uncertainty, consistency of a procedural constraint, or the absence of a prohibited downstream inference. A proposed test record should name the evaluative method and the actor authorized to speak for that evaluation; its result cannot by itself compel acceptance by all parties.

We can distinguish three outcomes for a preservation claim:

$$\text{Preserve}_{q,h}(r; p_q, E_q) \in \{\text{supported}, \text{refuted}, \text{undetermined}\}. \quad (48)$$

Here h fixes the represented configuration of evidence and review, not a universal present; *supported* means that the declared checks support the claim within a stated boundary, *refuted* that a stated counterexample defeats it, and *undetermined* that the evidence, access, or standards of comparison are insufficient or disputed. The statuses are evaluations of *claims under tests*; they are not intrinsic properties of a culture, language, institution or physical reality. A system can preserve two incompatible reviews as coexisting attributed records instead of manufacturing an institutionally authoritative consensus. A finite test set that contains no failure supports, at most, adequacy on the tested cases; it is not a proof of universal preservation over the declared admissible domain. Where a formally justified domain-wide theorem is available, its assumptions and proof may be attached as a different evidence type. Tests of equivalence should distinguish a confirmed *counterexample*, an *absence of observed counterexamples*, and an *unassessed* part of the domain. Appendix C adds a quantitative statement of unavoidable loss even when approximate reconstruction is allowed.

A deliberately lossy translation illustrates the distinction. Suppose A distinguishes protected ecological sites a_1 and unrestricted sites a_2 , while a proposed category in B represents both as b . If authorization, appropriate sampling procedures, or scientific risk differs between a_1 and a_2 , the many-to-one mapping cannot justify action on b without additional context. The correct infrastructural outcome may be a qualified scientific comparison and a withheld authorization, rather than a stronger semantic equivalence. The noninvertibility argument of Section 6 establishes a formal limit for selected transformations; the present requirement adds an operational duty to expose counterexamples, uncertain cases, and the authority boundary of the proposed use.

Four decisions must therefore be represented separately: whether a description or artifact is *technically exchangeable*, whether an inference or mediation is *scientifically warranted for the task*, whether the relevant actor has *competent authorization* for the proposed activity, and whether affected participants can exercise *effective objection or refusal*. These are not successively issued certificates of ethical legitimacy. Each may be contested, limited to a purpose, withdrawn where an applicable rule permits, or left unresolved. An AI mediator can produce candidates, identify cited sources, suggest counterexamples, and assist reviewers; it cannot convert a high-confidence semantic match into a legal entitlement or a binding representation of another community's epistemic practice.

The service should record the scope of disclosure as carefully as the claimed match. Sensitive source material may remain local while a contributor supplies a limited comparative test or a reviewed output. Where protected distinctions cannot be disclosed to every recipient, the parties may establish an independently accountable restricted review route; in other situations mediation may legitimately stop. *Refusal and unresolved difference are representable outcomes*, not necessarily system failures. The practical failure is to misreport these outcomes as successful full translation or to continue an unauthorized operation despite a known restriction.

10.5 Task-Relative Adequacy and the Limits of Composition

A mediation need not have a universal notion of success. For an explicitly identified task and a declared comparison space, it may be possible to test a narrow correspondence between outputs. Such testing

becomes mathematically well specified only when the compared quantities, observational supports, error measure, admissible domain, and relevant assumptions are fixed. If the task instead concerns institutional recognition, an obligation, or ecological standing, the parties may need an attributed record of reasons and procedures rather than an artificially numerical loss function. A common record envelope may hold both kinds of assessment without representing the second as a quantity in the first.

The fiberwise criterion of Proposition C.1 gives a limit on exact recovery after information loss. Its approximate counterpart, Proposition C.2, identifies a minimum possible uniform reconstruction error on finite real-valued examples. Both results assume a fixed source domain and target; neither says what a participant ought to value or whether a proposed use is authorized. For a chain of mediations, Proposition C.4 provides an error bound only when the intermediate output lies in the next validated domain and all three evaluations concern a justified common task. Context or authority is not closed under numerical composition.

For implementation, a translation warrant should therefore carry (i) the native and target record types, (ii) the declared task and comparison method, (iii) the validated input domain, (iv) known failures and unresolved cases, (v) the provenance of any composed mapping, and (vi) independent permissions and contestation routes. A successful numeric match cannot justify translating a protected ecological category into a public-access permission. Equally, refusal by an authorized steward to disclose a protected record is not evidence that the two scientific representations are incompatible. This separation keeps generalized translation experimentally investigable without collapsing relational meaning into a single universal metric.

10.6 Circulation of Heterogeneous Values and Relational Consequences

Circulation of knowledge may create and transform other values: scientific capability, ecological understanding, educational opportunity, collective recognition, commercial possibilities, medical applications, new rights claims, and potential harms. These do not share a natural common unit, and not every value is appropriately exchanged, owned, or discharged through a transaction. In particular, copying a genomic sequence does not relocate the living organism or reproduce its ecological dependencies; disseminating a model does not automatically transfer the capacity to validate it; payment for a data service does not necessarily resolve claims regarding collective authority or ecological responsibility. The relevant inquiry concerns *which relational configurations and capacities change*, for whom, through which operations, and under which rights and obligations.

For a declared circulation event e , a proposed record may be described by

$$\mathfrak{V}(e) = (s, t, x, \omega, \Delta\mathcal{R}, \mathcal{B}, \mathcal{O}, \mathcal{U}), \quad (49)$$

where s, t denote situated contributors and recipients, x the relevant artifact, action, or practice, ω the claimed type of circulation, $\Delta\mathcal{R}$ the documented or hypothesized changes to relational dependencies and productive capacities, $\mathcal{B}$ the represented benefits and burdens, $\mathcal{O}$ the implicated obligations, and $\mathcal{U}$ the uncertainty, exclusions, and contesting accounts. These fields are neither necessarily numeric nor necessarily commensurable; representations of $\Delta\mathcal{R}$ are limited by chosen observation boundaries. A record of a benefit claim must identify whose assessment supports it, rather than convert that claim into an uncontested property of the event.

To examine circulation, four operations should be kept separate: *recognition* of a contribution or value; *authorization* of particular uses or transfers; *realization* of some productive, scientific, or economic consequence; and *allocation or remedy* concerning the resulting benefits and harms. No operation entails the other three. This distinction prevents a provenance chain from functioning as a substitute for consent, a licence from being treated as evidence of ecological benefit, or a monetary valuation from exhausting heterogeneous forms of value. It also connects this section to the separate inquiry into heterogeneous value circulation and to the monetary analysis, without asserting that greater scientific information entails a reduction in monetary dependence [34, 37].

10.7 Value-Transformation Records and Dependency Boundaries

An engineering record should explicitly distinguish a *copy* of an informational product, *derivation* of a new product, *transfer or creation of a method*, *acquisition of productive capability*, *recognition of a contribution*, *authorization for a specified act*, and *realization of a benefit or harm*. The same research history may include all seven, but they have different observers, affected parties, evidentiary standards, and consequences. For example, dissemination of a workflow can increase another institution's technical capacity only if the necessary instruments, expertise, input access and material resources are also obtainable. Conversely, a recipient can obtain a commercial advantage from publicly accessible information without gaining or providing any rights over the contributing community's protected material.

An event graph may represent declared transformations and accountable dependencies. Let $\mathcal{E}_h$ denote known events under representation h , and let $e \rightsquigarrow_h e'$ indicate a reported dependency relevant to a stated purpose. Define a scoped reachability set $\text{Reach}_h(e) = \{e' \in \mathcal{E}_h : e \rightsquigarrow_h^* e'\}$. For a declared correction δ and review task q , let $I_{q,h}(\delta, e') \in \{\text{affected}, \text{unaffected}, \text{unresolved}\}$ be an attributed impact screening, supported by stated evidence. The associated *candidate* review set is

$$\text{Review}_{q,h}(e; \delta) = \{e' \in \text{Reach}_h(e) : I_{q,h}(\delta, e') \in \{\text{affected}, \text{unresolved}\}\}. \quad (50)$$

The reflexive-transitive closure refers to *known, recorded* dependencies within a declared search boundary. An *unresolved* impact result, including a not-yet-screened reached item, is retained for review rather than silently treated as harmless; an *unaffected* result requires a documented rationale, not a missing response. Neither a graph query nor the screening proves that the listed dependents actually changed. No complete graph of world-wide copying, hidden influence, or downstream practice is assumed. In particular, $\text{Review}_{q,h}$ is complete only relative to the *declared* graph and screening rules; it cannot establish exhaustiveness over real uses of the original observation. A notification and review workflow may inspect identifiable derivatives, rerun affected computations, qualify published claims, or refer disputed uses to competent institutions. Its success should be reported through coverage, response and remaining uncertainty, rather than the unsupported assertion that all consequences of a source error have been undone.

The event record should also distinguish *asserted* from *realized* consequences. A patent application, an estimated economic value, an expected therapeutic benefit, and an observed medical outcome occupy different evidential and institutional positions. Similarly, community recognition, consent to one sample collection, permission to publish a derivative, and an enforceable benefit-sharing obligation may have different scopes and authorities. Provenance may support analysis of these differences; it neither establishes the validity of disputed claims nor provides a universal unit in which ecological integrity, recognition, monetary revenue, and public scientific capability can be summed. Accounts should permit multiple attributed evaluations and should identify missing or unrepresented interests.

10.8 The Ocean as an Integrated Physical and Ecological Testbed

A genuinely global knowledge infrastructure must also connect to its material means of observation. Marine research makes this connection unusually visible because sensing, underwater navigation, computation, energy, communications, ecological context and legal competence are closely coupled. Autonomous underwater vehicles, remotely operated vehicles, profiling floats, gliders, surface relays, fixed observatories and research vessels support different sampling and operational regimes. Their scientific results depend on mission design, calibration, deployment and recovery histories, pressure and corrosion constraints, navigation error, communications availability, and post-processing. The resulting federation should represent the measurement chain and its limitations rather than treat "ocean data" as a homogeneous input. Existing Argo and GOOS practices supply important domain-specific precedents, not a template that settles every future marine service [55, 58, 30].

An adaptive marine robot could use local model uncertainty to propose a sampling location or measurement, yet this recommendation must be evaluated against mission safety, scientific sampling bias,

Table 16: Coupled marine infrastructure elements and proposed epistemic and ecological interfaces. Rows distinguish engineering needs; no row asserts field validation or permission for a new deployment.

Material subsystem	Records and coordination interfaces	Main research or governance constraint
Robotic and fixed observation	Missions, calibration, sampling effort, instrument and deployment histories	Observability bias, reliability, habitats and safe recovery
Energy and charging	Power source, storage, docking, energy use and maintenance assumptions	Reliability, upstream materials, ecological and lifecycle costs
Communications	Acoustic, optical, cable or surface-relay capabilities; outage and synchronization logs	Variable bandwidth, intermittency, missing observations and security
Marine biological inquiry	Sampling provenance, taxonomic uncertainty, sequence or ecological representations	Disturbance, sensitive sites, contested reuse and biodiversity responsibilities
Distributed computing	Onboard inference, shore-side processing, model dependencies and energy accounts	Reproducibility, ownership, capacity asymmetry and environmental footprint
Experimental subsea facilities	Thermal balance, pressure enclosure, service/retrieval plan and baseline monitoring	Uncertain comparative benefit, repair burden, marine heat and permitting

ecosystem effects, and any binding authorization. More measurements do not necessarily improve ecological understanding if they repeatedly sample convenient regimes or alter sensitive habitats. A closed-loop design should retain competing scientific hypotheses, the reasons for selecting observations, the measurements not taken, and human review of consequential interventions. Biological or ecological records also warrant differentiated disclosure: some precise localities or community-linked practices should not become public merely because a research instrument collected data there.

Energy provision is part of the epistemic system because it shapes which observations and analyses are feasible. Surface solar, offshore supply connections, batteries, vehicle docking, and location-dependent marine energy options entail different reliability and environmental profiles. Similarly, onboard inference can reduce transmission demands but may make later independent audit harder if raw data are not retained or if model versions and uncertainty are unavailable. Designs should compare feasible energy, maintenance, transport, computation and communication arrangements under explicit lifecycle boundaries, including recovered materials and eventual decommissioning.

Subsea computing deserves separate treatment. Submerged computing enclosures and seawater heat exchange are technically investigable, but potential advantages depend on electricity sources, network links, thermal dispersal, maintainability, failure recovery, and site conditions. An underwater data centre should thus be considered an experimental option with unresolved comparative engineering and ecological evidence, not a necessary node in a first-generation knowledge federation. A software pilot can instead use existing onshore or cloud resources and actual public marine records; new subsea deployment requires site-specific environmental assessment, safety and regulatory review. The ocean is an evolving biological and ecological setting containing affected human interests, not an empty location into which infrastructure may be placed without obligation.

10.9 A Cross-Border Marine Knowledge Mediation Case

Consider an explicitly fictional marine research arrangement involving an observation institution A , a laboratory B , locally situated ecological knowledge holders C , and a relevant public authority J . Participants may belong to two countries, but each country contains multiple competing epistemic and institutional positions. An autonomous ocean platform operated by A produces environmental observations and identifies a microorganism of possible scientific interest. A research team at B combines lawful

samples or records with its own methods to investigate a biochemical mechanism. Participants at *C* may hold ecological observations or relationships concerning the environment that are scientifically relevant but expressed through different categories and practices. The hypothetical case does not establish that a candidate compound will be medically useful, that all biological data may lawfully circulate, or that any specific person or community possesses a legal entitlement.

At the first interface, *A* and *B* require calibrated measurement descriptions, record identities, sampling limits, and reproducible analyses. At the second, some terms used by *C* and *B* may lack direct counterparts; an elicitation and translation protocol should record differences rather than forcing a dominant classification. At the third, claims concerning biological samples, sequence information, confidential knowledge, patentable inventions, collective rights, or benefit-sharing demand a *jurisdiction- and activity-specific legal assessment*. At the fourth, new scientific or commercial capacities may become concentrated in *B* despite earlier contributions by *A* and *C*. Finally, an intervention justified as conservation or innovation might alter the original ecosystem, requiring ecological monitoring and possible remedy.

A successful infrastructural test would not require a unanimous monetary valuation of this episode. It would require a reconstructible record of the authorized observation and scientific inferences; explicit statements of what the translation did and did not preserve; competent legal and institutional review where needed; meaningful channels for affected participants to challenge uses; and enough independent access to assess claims without compelling disclosure of protected information. Its adverse cases would include a persuasive but false semantic match, a derivative AI output that discloses protected knowledge, a purported benefit-sharing condition that lacks legally competent authorization, an undisclosed use of environmentally harmful sampling, and reproduction of a scientific result that remains impossible for the contributing node. The laws applicable to these questions vary by location, material, actors, and activities; this paper does not assign a comprehensive legal status to the fictional case.

10.10 Implementable Services, Protocol Boundaries, and Normative Review

Table 17 turns the research proposal into a set of prospective services and bounded acceptance tests. The services may be operated by multiple institutions and need not be held in one database, blockchain, AI agent, or globally authoritative registry. Federated catalogues can draw on DCAT 3; provenance on PROV-O; observation metadata on SOSA-SSN and relevant domain standards; local conceptual alignments on SKOS. ODRL supplies an expression model for certain permissions, prohibitions, and duties. Such syntactic and information-model compatibility does not establish the correctness of scientific inferences, the validity of a contract, the status of customary law, or the authority of the party expressing a condition [74, 71, 75, 70, 73].

An implementable *mediation dossier* should at minimum contain a stable local identifier; the bounded task and participating contexts; versioned pointers to origin artifacts and models; claimed correspondences and counterexamples; explicit uncertainty and protected fields; a separately verified authority or consent status where applicable; a record of permitted downstream operations; and links to correction, dispute, and redress procedures. Some entries may be withheld from public inspection for legitimate reasons, in which case the federation should consider whether purpose-limited external auditing is possible. A cryptographic signature can authenticate a record or actor under its trust assumptions, but cannot adjudicate scientific validity or transnational legal competence. Community-governed protocols such as the Traditional Knowledge and Biocultural Labels developed through Local Contexts offer a practical precedent for communicating community-specific expectations about provenance, use and access. Local Contexts distinguishes provenance, protocol and permission labels; such labels do not by themselves confer a statutory right on a recipient or make every proposed use lawful. Their meaning and legitimate application depend on the relevant communities and cannot be assigned unilaterally by an automated mediator [45, 6].

The architecture should preserve the separation already developed in Section 8: a participant's acceptance, formal rule validity, legally effective enactment, and ethical legitimacy can diverge. The resulting

Table 17: Candidate engineering and normative services for shared knowledge production, generalized translation, and heterogeneous value circulation. The tests are proposed design criteria, not completed implementations.

Service	Principal deliverable	Failure to expose in a trial
Common research workspace	Versioned questions, methods, evidence and revisions	Reusable result with no traceable research process
Instrument and resource federation	Calibrated task requests, budgets and permissions	Nominal membership without usable instruments or compute
Discovery and circulation	Distributed catalogs, controlled derivatives, correction links	Discoverable data with no warranted reuse or correction
Generative translation	Partial mappings, task tests and explicit loss records	Fluent correspondence hiding a critical incompatibility
Value and contribution relations	Plural benefit, dependence and obligation histories	Payment or attribution mistaken for complete recognition
Normative authorization	Applicable-rule dossier, reviewers and decision histories	Automated policy assertion treated as legal authority
Contestation and remedy	Objection, review, revision, notification and appeal route	Authentic but harmful output without effective response
Ecological and material accounting	Mission energy, maintenance, habitat and lifecycle record	Apparent local benefit with displaced ecological costs

workflow is therefore a *review-support system*, not an autonomous court, legislature, or universal moral calculator. AI can assist in identifying potentially applicable sources, alternative interpretations, and inconsistencies, but consequential legal classification or authorization requires accountable institutional review, especially where legal traditions, languages, and jurisdictions differ. A dispute should be capable of producing revision, reservation, restricted circulation, alternative interpretation, or a documented unresolved state; it need not terminate in compelled consensus.

To connect these services to the federation rather than operate an isolated translation platform, a circulation request should reference its originating observation or scientific artifact, the declared research operation, the bounded mediation dossier in Equation (47), and any separately recorded decision about the particular downstream use. Equation (18) acts as an interface discipline: technical acceptance, interpretive warrant, competence of the relevant authority, and ability to challenge are retained as distinguishable attributed assessments. The service must support partial review, multiple reviewers, protected evidence, conflicting responses, and nonresponse. The profile supplies no universal adjudicator, and does not eliminate the need for participants to determine the lawful and ethical status of the underlying acts.

In a minimal deployment, a request to circulate a translated marine habitat result could be marked technically valid and scientifically supported for a regional summary, while permission to release precise site coordinates remains unresolved and a stewardship objection is pending. The public endpoint should consequently publish only the material covered by an independently justified disclosure decision. An unrestricted public result must not be assembled by treating absent evidence of a restriction as affirmative permission. This test simultaneously exercises the scientific bridge of Section 6, the authority distinctions of Section 8, and the dependency records of Section 10.7.

10.11 Authority, Stewardship, and Long-Term Operation

The service contracts require institutional counterparts. At minimum, a proposed federation should identify the custodians and scientific stewards of each resource; the accountable operator of each computing or robotic service; the persons or bodies competent to decide a specified request under applicable ar-

rangements; the scientific and legal review channels; and the affected parties entitled to consultation or challenge under applicable rules. A registry may record *claims* concerning these roles without deciding disputed competence. Similarly, cryptographic authentication, contractual representation, scientific peer review and lawful authorization answer different questions. A technical operator should not acquire normative authority simply by hosting the mediation service.

Long-term operation creates additional design obligations. Costs include maintenance, calibration, storage, software and cybersecurity updates, translation review, legal assistance, training, low-bandwidth access, physical retrieval and ecological monitoring. Governance has to specify how scarce services are allocated, how small nodes can enter and leave, whether essential methods and data can be exported, how an incompatible protocol change is handled, and how a challenge reaches an actor able to act upon it. A public claim of capacity-sharing should be accompanied by evidence of completed resource allocations and independently reproducible work, not only resource catalogues and promised access. Funding, hardware supply and cloud-service concentration can make a nominally decentralized federation operationally dependent on a small number of suppliers.

For consequential translation or downstream use, a review-support workflow can distinguish a submitted proposal, an evidence-based interpretation, an identified institutional decision, an enacted action, a challenged outcome, and any implemented correction or remedy. These stages are distinguishable observations, not a mandatory linear sequence and not a universal model of judicial or legislative procedure. A refusal, incompatible decision, nonresponse, or lack of competent forum should remain visible as a bounded state rather than be silently repaired by AI inference. Authorization should be scoped to an action, actor, materials, jurisdiction and relevant conditions; a valid decision for one use does not automatically authorize training a model, transferring a sample, publishing precise locations or creating a commercial derivative.

10.12 Legal Knowledge, Technological Change, and Competent Deliberation

An evolving technological environment also requires infrastructure for studying the *development of applicable norms*, rather than a static catalogue of machine-readable prohibitions. A legal knowledge service should retain authoritative source versions, promulgation and applicability information, jurisdictional and institutional scope, relevant interpretive materials, identified amendments or repeals, and reasoned challenges to a proposed application. It should distinguish a documentary record of a rule from an attributed interpretation of its meaning, a party's claimed entitlement, a decision by a competent institution, and evidence that a decision was enacted or a remedy delivered. Lawful procedures for obtaining and contesting these records remain dependent on the relevant legal order.

Highly capable AI and autonomous robotic systems complicate normative classification. New research practices may combine observation, biological sampling, model training, delegated operations, international computation, and downstream commercial or public uses that different bodies govern under different mandates. A generalized translation service can assist by identifying similarities and differences in terms, evidential burdens, procedures, recognized authorities and protected interests across legal contexts. Yet a technically coherent crosswalk between two legal classifications cannot transfer one jurisdiction's institutional competence into another. Nor does the generation of a previously unanticipated technical capacity itself confer a right to deploy it. Proposed acts should be decomposed into their legally consequential operations, with separately documented authority and uncertainties for each.

A research prototype might therefore offer four bounded functions: *source retrieval and version comparison*; *issue and applicable-authority identification*; *conflict, omission and alternative-interpretation annotation*; and *referral to competent review with a record of the decision and any later challenge*. None of these functions automatically resolves indeterminate jurisdiction, conflicts of law, disputed representation, or questions of legitimacy. The programme should test false-positive and false-negative rule matches, outdated legal versions, fabricated citations, changes in actual institutional competence, and situations in which a technically permitted operation nonetheless warrants an ethical or ecological objection. AI may help with scale and access while its confidence and outputs remain distinct from legal

authority and publicly accountable deliberation.

The institutional dimension includes procedures for technological learning. Regulatory experiments, standards review, community consultation and documented incidents may contribute evidence relevant to changing a rule, but no such process guarantees improvement or a single direction of evolution. Infrastructure can preserve the provenance of proposals, competing interpretations, dissent, decisions and revisions; authority to alter a rule remains with the applicable bodies and practices. This separation allows generative knowledge systems to inform normative change without assigning legislative, judicial or collective-governance powers to a translation algorithm.

10.13 A Release-Gated Federation Pilot

The preceding service description permits a bounded implementation claim: a prototype could coordinate the *processing* of a request while retaining its scientific and institutional review boundaries. For evaluation purposes, consider two independently operated research nodes, a designated mediation service, a correction index, a restricted evidence vault and a human-operated review channel. The nodes hold different, deliberately constructed ecological classifications and associated model records. The first implementation uses fictional actors and publicly reusable observations; its objective is to test whether a downstream action remains attributable, reversible where technically possible, and responsive to appropriately scoped objections. A shared identity service and a record catalogue would be implementation conveniences, not proof of a common ontology or of legitimate global jurisdiction.

For a proposed operation, a release-gate record should specify its identifier, actor, task, source version, scope and destination. It should also contain attributed scientific assessments, evidence of relevant institutional decisions, challenge and withdrawal states, and the proposed public or restricted output fields. A request to publish a coarse habitat summary, a request to release exact coordinates, and a request to train a model on the same underlying record are different operations. Authorizations, where independently established, must be assessed for the particular actor, purpose, material, duration and operative context. A declaration produced by an AI assistant remains a *claim about permission* until assessed through competent processes; the service cannot certify that the claimed institution is competent merely by observing a signature or an apparently well-formed rule statement.

A test interface may return release, hold-for-review, or reject-for-declared-use. These are *provisional processing decisions* in a constructed pilot. Release means that the specified pilot gate has found adequate recorded grounds for its limited output, while remaining subject to the applicable rules and accountable operators. Hold-for-review includes absent authority evidence, a plausible but untested translation, a pending objection and incomplete dependency inspection. Reject-for-declared-use includes a proven mismatch for the declared task, an expressly prohibited output, a scope conflict, or a withdrawn source whose proposed reuse depends upon its obsolete version. The gate must not infer permission from nonresponse, interpret a test certificate as a legal ruling, or replace unresolved results with invented certainty. In an appropriately designed pilot, a public endpoint must also omit protected fields even when a restricted research task is scientifically supported.

The sequence in Table 18 is a hierarchy of *evidential demands*, not an asserted universal sequence of technological or institutional development. A project can remain at one configuration; some studies require additional forms of inquiry and some forms of evidence cannot safely be generated. A real-world governance review should be constituted under the relevant law and institution, rather than by promotion of a test-harness result. The machine-readable pilot cases in the accompanying source package only exercise a locally defined decision procedure over invented statuses.

10.14 Comparative Translation, Correction, and Capacity Trials

For bounded translation tasks, the central comparison is between making more confident correspondences and making *better qualified* correspondences. An initial preregistered design could compare (i) simple term or schema alignment, (ii) alignment supplemented by elicited generative hypotheses, counterexamples and documented non-equivalences, and (iii) the second approach with accountable review

Table 18: Prospective pilot gates and evidence required before a claim can pass from one research configuration to another. Passing is necessary for the specified study transition, not sufficient for a lawful or ethically acceptable real-world deployment.

Evaluation configuration	Required evidence and independently accountable function	Hold or stop condition
Synthetic record trial	Frozen schemas, fictional rule cases, versioned source graph, expected review states; independent test operator	False public release, fabricated authority or unrecorded dependency failure
Cross-node software sandbox	Two separately administered nodes, documented adversarial inputs, outage and recovery logs, authenticated review decisions	Uncontrolled sensitive disclosure, unexplained semantic loss or no actionable correction path
Domain-scientific trial	Appropriately licensed observations, calibration and independent holdouts, scientific reviewers and reproducibility resources	Systematic miscalibration, unjustified equivalence, unusable reproduction or obscured sampling bias
Governed material operation	Competent permits and consents as applicable, safety case, ecological baseline, consultation and independent oversight	Missing authorization, unacceptable safety risk, unknown sensitive impact without suitable protections

and correction interfaces. Each arm should receive matched tasks and a defined opportunity to decline correspondence. Reviewers should evaluate an arm’s decisions from independent reference descriptions, with uncertainty over the reference itself recorded. Where human participants are involved, ordering, reviewer familiarity, translation direction and learning effects must be controlled or openly treated as limitations. A higher mapping count does not establish greater epistemic quality.

For a finite, eligible test set Q under a declared independent adjudication procedure, let U denote cases with a task-critical correspondence known to be unsupported and let L denote cases actually released by the *pilot* gate. The synthetic unsafe-release rate is

$$\text{UnsafeRelease}_Q = \frac{|U \cap L|}{|U|}, \quad |U| > 0. \quad (51)$$

The reference status and its limits must be documented; a disputed or unknowable reference cannot be silently placed in U . In parallel, a *qualified-use yield* should be reported on cases where a supported release could in principle be justified, and the rate of review deferrals should be disclosed separately. Otherwise a system that refuses every request might appear perfectly safe while providing no knowledge-producing capacity. For legally meaningful decisions, the corresponding safety findings must be established through domain- and jurisdiction-appropriate review; Equation (51) is a measurement definition for a controlled study, not a legal compliance index.

A correction trial should inject a documented error into a source and compare the *recorded reachable dependencies*, the subset actually screened, recipients notified, computational revisions attempted, changes independently verified and unresolved cases. The denominator of a notification rate is the specified population of *known contactable recipients*; a missing or unknowable recipient is an explicit limit, not an affirmative success. A negative effect found in a descendant is evidence of that descendant’s sensitivity to the correction under the tested conditions, rather than proof of the entire societal influence of the source. The operational question is whether a bounded challenge can lead to an auditable change or a reasoned refusal.

The productive-capacity trial asks whether less-resourced contributors can conduct an independently checkable analysis, not simply download its final artifact. Eligible participants may receive different combinations of compute access, instrument time, methodological guidance and controlled processing privileges. Predefined outcomes include task completion under scientifically comparable conditions,

reproducibility, assistance costs, time, unsuccessful applications and sustained ability to challenge a result. The evaluation must document the selection of eligible participants and avoid withholding essential services or legally protected rights to produce an attractive control group. Differences between access permissions, effective allocations and realized scientific practice should be maintained in the final analysis.

10.15 Material Feasibility, Ecological Safeguards, and Institutional Stopping Rules

The marine example permits a disciplined choice among engineering configurations. The baseline is reuse of appropriately licensed observation records with conventional terrestrial computation. A second, more demanding option adds a simulated vehicle mission, a power and communications model, and uncertainty in sensor placement. A third option introduces remotely supervised physical measurement under a documented safety and environmental review. Autonomous habitat intervention, genetic sampling with disputed rights, or experimental seabed computing are separate proposals with distinct permissions and ecological effects; they do not follow automatically from success in earlier software studies.

A marine evaluation dossier should record its scientific objective, feasible alternatives, functional units, operating and maintenance boundary, power and communication assumptions, retrieval and decommissioning plan, habitats and communities potentially affected, and assumptions about ecological response. It should use separate environmental indicators as in Equation (103); a favorable local energy-intensity result does not compensate by definition for biodiversity effects or a displaced supply-chain burden. A research plan also needs a baseline including uncertainty, qualified ecological review, and triggers for stopping or revising operations if the consent, safety or ecological conditions supporting the plan change.

Table 19: Illustrative research choices and the evidence needed to examine them; this table makes no feasibility or comparative-benefit finding.

Option	Measurable engineering quantities	Additional review boundary
Existing data and onshore compute	Processing demand, reproducibility, sampling limits and access constraints	Dataset rights, uncertainty and downstream disclosure
Vehicle simulation	Energy budget, mission loss probability, latency, missed sampling strata	Simulation-to-field limits; no real habitat effects measured
Supervised marine sensing	Calibration, equipment reliability, recoverability, spatial sampling and maintenance	Site-specific authorization, habitat assessment and operational safety
Experimental subsea computing	Total electricity, thermal dispersion, retrieval, repair, network dependency and lifecycle materials	Site-specific environmental assessment and competent review before physical trial

Institutional validity cannot be supplied by an ecological indicator or a technically coherent policy file. For each material or data-transfer operation, responsible persons should document the potentially applicable authorities, the consulted sources and their versions, jurisdictional limits, reasons for alternative legal interpretations, decisions actually issued, and possible paths for review and correction. The absence of an identified competent forum is an unresolved institutional problem that may prohibit progression; it is not an invitation for software to invent an authority. Affected participants' ability to decline involvement, maintain protected information and challenge proposed uses must be assessed through the appropriate procedures. Conditions for stopping a study should be set *before* consequential operations and revised when new evidence or binding decisions warrant it, rather than treated as adjustable rewards for favorable experimental results.

10.16 Practical Deployment, Evaluation, and Research Boundaries

A limited first prototype would integrate a public marine or meteorological dataset, a versioned research workflow, two deliberately incompatible interpretive schemas, a synthetic set of governance constraints, a mediation record, and a correction propagation service. The demonstration should use clearly fictional institutions, permissions, and contribution claims; it should neither appropriate protected community knowledge nor imply the existence of an enacted cross-border agreement. A second trial could compare human-only mediation with human-supervised AI assistance on declared tasks. Participants would judge scientific adequacy and interpretive loss independently of technical completeness, while qualified legal reviewers would assess jurisdiction-specific normative claims. Reproduction capacity, environmental effects, and distribution of benefits require their own measures rather than a single composite efficiency score.

Tests should include deliberately ambiguous mappings, incompatible explanatory models, conflicting permissions, a withdrawn or corrected source record, a community objection, an unavailable central service, and a high-quality translation that does *not* confer authorization. For each case, the trial would record whether an unsupported equivalence was exposed; whether downstream uses could be identified; whether restrictions were respected; whether a challenge could reach an appropriate reviewer; and whether a resource-constrained participant could reproduce or contest the result. A system may pass a schema test while failing any of these epistemic, legal, or public-capacity tests. In particular, increasing the number of successfully mapped items can correlate with a rise in overconfident or unauthorized mediation.

Ocean robotics, marine genomic research, coastal observations, energy provision, and possible sub-sea computation supply rich engineering contexts but have distinct maturity and risk profiles. Readily accessible, lawful datasets and existing computing services are sufficient for an initial prototype. New underwater robots, persistent subsea power, autonomous ecological interventions, or subsea data centres require independent engineering, environmental-impact, safety, and regulatory evaluations. The research therefore begins with reversible, bounded demonstrations rather than presupposing deployment of speculative infrastructure or treating the ocean as unoccupied industrial space.

Three successive research configurations can make the proposal testable without presuming an inevitable development sequence. An *offline synthetic configuration* can demonstrate the validity and limits of records, correction-reach queries, deliberately invalid mappings and dispute routing. A *cross-node scientific configuration* can evaluate two independently operated research workspaces using openly reusable marine records, incomplete connectivity, reproducibility requests and supervised AI mediation. A *governed field configuration*, contingent on competent cooperation and independent review, can investigate actual instrument operations, restricted records, participating institutions and ecological outcomes. Each configuration has separate entry criteria and failure conditions; successful synthetic tests do not confer permission or scientific validity on later configurations. The companion study can specify performance budgets, adversarial test protocols, institutional consent processes, and experimental uncertainty at the level appropriate to each case.

This section enlarges the meaning of a knowledge federation: its purpose is to make the *conditions of knowing and creating* more shareable, traceable, contestable, and capable of development, while respecting contexts in which circulation must be restricted or refused. It does not claim that all meanings can be translated, all values recognized by one evaluator, or all obligations resolved through interoperable records. The complementary paper proposed in Section 13 should investigate the construction, feasibility, and limitations of these services as its own object. The scenarios in the following section retain their role as falsification-oriented tests for the broader global observation architecture.

11 Hypothetical Design Scenarios and Evaluation

The preceding sections specified the epistemic and institutional conditions of a proposed global observation infrastructure. Those conditions are meaningful only insofar as they admit tests that could reveal

where the proposed architecture fails. The objective of this section is therefore to construct a set of *stipulated design scenarios*, each with a well-defined observation boundary, a declared task, and an adverse case. None describes an implemented federation, a measured international programme, or an evaluation of an identifiable institution. Their purpose is to make the architectural proposal answerable to empirical, technical, and institutional evidence.

Several methodological distinctions guide the scenarios. A common catalogue entry does not establish scientific comparability; the same published product does not imply equal capacity to reproduce it; a locally updated permission record does not by itself retract previously copied data; and scientific confidence can be misplaced even when every contributing sensor has passed its respective local check. Conversely, infrastructural autonomy and institutional plurality are compatible with a degree of centralized service provision. Accordingly, the tests compare documented *properties and consequences* rather than treating decentralization, public access, or model agreement as intrinsic success criteria. Institutional examples such as WIS2, Argo, and international early-warning cooperation inform plausible interface requirements, but the synthetic cases below add further, unimplemented components [68, 2, 59].

Let a scenario be a bounded design dossier

$$\mathfrak{S}_j = (\mathcal{N}_j, \mathcal{D}_j, \mathcal{M}_j, \mathcal{G}_j, \mathcal{H}_j, \mathcal{K}_j), \quad (52)$$

where $\mathcal{N}_j$ denotes participating observation and computational nodes, $\mathcal{D}_j$ the available records and their quality histories, $\mathcal{M}_j$ the models and interpretive procedures, $\mathcal{G}_j$ the institutional permissions and responsibilities, $\mathcal{H}_j$ a partially ordered collection of documented events, and $\mathcal{K}_j$ the explicit evaluation tasks and criteria. In Equation (52), a history is not assumed to lie on a universal physical or social timeline: causal, documentary, and institutional orderings must be specified for each task. The dossier also records exclusions, including external data sources and actors whose decisions may affect the outcome. A positive result within a chosen boundary is not a claim about the global system beyond it.

11.1 A Cross-Border Meteorological and Ocean Data Federation

Consider three hypothetical participants: a national meteorological service N_M , an independently governed ocean observing centre N_O , and a regional warning institution N_W . A fourth node N_R operates a metadata catalogue and notification relay without holding every underlying record. N_M supplies atmospheric observations and forecasts; N_O supplies temperature and salinity profiles and relevant quality flags; N_W combines authorised products for a coastal-hazard question. This is a deliberately limited demonstration of complementary observational capabilities. It does not presume that a particular ocean variable can directly determine a public warning, or that a global data-policy category overrides an institution's applicable duties.

The synthetic test begins when N_W requests a specified product covering a named coastal domain and observational support. It attempts to discover the contributing records, identify spatial and temporal supports, retrieve or lawfully query them, and reproduce the derived product with its declared model version. Discovery success, retrieval success, scientific fitness, and authorization success are recorded separately. A federation could score perfectly at notification delivery while failing to supply calibrated ocean data for the task. A second failure occurs if ocean profiles are discoverable but their depth support or delayed quality-control revisions are ignored, making the derived result scientifically misleading. Existing WIS2 services distinguish discovery, notification, cache, and provider access; Argo distinguishes timely release from scientifically reviewed delayed-mode processing. These precedents motivate a layered test rather than a claim of direct operational integration [64, 1, 55].

A hypothetical workflow contains the following registered events: publication of a dataset and discovery metadata; dissemination of a notification; an authorized retrieval or computational request; assignment of a calibration and model version; production of the warning-oriented result; and, where warranted, a scientific or institutional revision. The event dependencies must be checkable even when publication, retrieval, and warning institutions maintain different local clocks. A correction originating at N_O should produce a machine-readable linkage to the affected products at N_W , not merely overwrite the original

profile. The resulting provenance graph is therefore part of the scientific output, not administrative decoration [71].

For this scenario, assess *product completion* independently of its *reproducibility and operational consequence*. Required evidence includes retrievable version identifiers, valid access decisions, unit and support compatibility, correction propagation, and review of the derived warning against suitable reference outcomes. Warning dissemination and public response have additional institutional dependencies that a data federation alone cannot guarantee. We should consequently compare failure recovery and correction latency as well as successful initial requests. A technically available prediction that cannot be independently challenged, corrected, or interpreted by the responsible warning body remains an incomplete demonstration of shared knowledge infrastructure.

11.2 Sensor Drift and Version Correction

The second scenario examines an apparently ordinary sensor update that creates a difficult cross-node epistemic question. Let a sensor at event e generate

$$y_e = H_e(x_e) + b_e + \epsilon_e, \quad (53)$$

where x_e is an environmental state over the declared measurement support, H_e its observation operator, b_e a systematic instrument bias, and ϵ_e an observational error term. The task is not simply to estimate x_e ; it is also to decide when a published observation should be marked, corrected, quarantined, or retained as an earlier version, and what downstream scientific products must be reconsidered.

For a fully specified toy example, suppose an observation operator equals the identity and the true test value is stipulated as 10 arbitrary units. An earlier measurement reports 10; a later measurement reports 12. The apparent increase may represent environmental change, instrumental drift, a change of support, or a combination. With only the two sensor readings, there is no principled basis for assigning the difference to any one mechanism. In an intervention arm, an independently calibrated reference observation reports 10 on the later support, and investigation documents a +2 unit instrument offset. The responsible institution then publishes a corrected value while preserving the original record, the evidence for correction, and dependencies of already produced results. These numbers are invented to test the workflow, not observed environmental measurements.

Proposition 11.1 (Unidentified state and additive drift from unanchored observations). *Under Equation (53) with H_e the identity, suppose a record contains only $y_e = x_e + b_e + \epsilon_e$ and supplies no independent constraint linking x_e or b_e to other measurements. Then the pair (x_e, b_e) is not uniquely identifiable from y_e , even if ϵ_e is known to be zero.*

Proof. For any compatible constant c , the alternative pair $(x_e + c, b_e - c)$ gives the same sum $x_e + b_e$. Distinct choices of c therefore yield identical observations. Additional calibration, cross-instrument evidence, physically justified constraints, or other identifying information is necessary to distinguish state from drift. $\square$

The proposition is a structural statement about the specified toy model, not a general impossibility of detecting drift. Real sensors may contain multiple observables, repeated standards, temporal constraints, or independently validated physical relationships that support identification. The proposed system must represent which of these sources justify a correction and who has authority to issue it. A node that silently replaces 12 with 10 would destroy evidence of the correction process. A node that preserves every version but leaves derived warnings unchanged would retain provenance while failing to support scientific repair. Good performance requires both: an auditable revision history and demonstrable examination of affected downstream claims.

The test should intentionally introduce two adverse cases. In the first, the reference instrument is unavailable or disagrees with the first institution; the system should record unresolved interpretations rather than force consensus. In the second, a corrected source observation has already entered multiple

models, some of which have been reused by outside institutions. Complete correction propagation may be infeasible where downstream dependencies are unknown. That limitation should appear in the report as an observed coverage bound, not be hidden by calling the source correction globally complete. These requirements draw on the WMO instrument guide, whose October 2026 edition is expressly provisional and was scheduled for approval consideration in November 2026; Argo quality-control practice; and the provenance concepts introduced in Section 4 [67, 56, 71].

11.3 Semantic Mismatch and Model Disagreement

In the third case, two institutions publish quantities with superficially similar labels but different scientific meanings. For example, an ocean sensor may report dissolved oxygen in a concentration unit while another product reports oxygen saturation, which also depends on environmental conditions and a specified conversion model. A syntactic service can make both fields retrievable, and a semantic service may identify a relation between the concepts, but neither establishes their interchangeability without the requisite contextual quantities, conventions, and uncertainty. This is a hypothetical integration task, not an assertion about the metadata quality of actual programmes. SOSA/SSN, CF conventions, and vocabulary standards offer useful description mechanisms but do not themselves supply all physical translation rules [75, 7, 70].

Represent a candidate translation between source and target interpretive systems by a partial, task-indexed map T_q and a target quantity ϕ_q . A test must specify the domain on which T_q is meaningful, the source and target observational supports, and acceptable target error. We can record its empirical discrepancy as

$$\delta_q(T_q; \mathcal{V}) = \sup_{v \in \mathcal{V}} d_q\left(\phi_q^{\text{target}}(v), \widehat{\phi}_q^{\text{target}}(T_q(v))\right), \quad (54)$$

where $\mathcal{V}$ is a declared validation collection and d_q a task-appropriate discrepancy. Equation (54) is a design metric, not a guarantee of generalization beyond $\mathcal{V}$. If the conversion is undefined for missing contextual variables, those records should be identified as unsupported, not converted by inventing parameters or assigning falsely precise outputs. Where a physical translation is not invertible, original supports and loss declarations must remain accessible to the permitted reviewers.

A second layer concerns model disagreement despite valid semantic alignment. Stipulate two models M_1 and M_2 that fit a shared calibration subset yet generate different conditional predictions for a held-out region or environmental regime. The evaluation then asks whether the federation exposes both predictions, their training and validation supports, their assumption sets, and the differences relevant to a user’s task. It should not select a single interpretation merely because one model is more commonly installed or computationally convenient. Disagreement can arise from model form, unresolved parameters, missing ecological processes, or legitimate methodological choices. The tests must attempt independent validation and report coverage, systematic errors, and uncertainty under each scenario, building on the distinction between observation and inference developed in Section 7 [22, 33].

We should include an intentionally deceptive case: two metadata catalogues use identical variable names and units, yet one product is a station observation and the other a spatially averaged model estimate. A name-and-unit matcher will declare success; a task-sensitive validator should identify their different measurement supports, processing, and epistemic status. The corresponding failure is not merely a software defect. It is an epistemic error that may systematically advantage institutions whose interpretive defaults are encoded in the federation’s most visible services. A defensible architecture therefore permits versioned mappings, alternative interpretations, disputed claims, and correction channels rather than making one global semantic schema the ultimate arbiter of meaning.

11.4 Divergent Interpretations and a Shared Normative Protocol

Consider a fictional federation in which a marine observatory, a coastal stewardship institution, and an emergency-warning agency receive access to the same verified coastal temperature and species-occurrence record. Their scientific and practical interpretations differ. The observatory prioritizes

independently reproducible location-specific research; the stewardship institution judges public coordinates potentially harmful; and the warning agency needs sufficiently granular information for a specific emergency task. These positions are a stipulated scenario rather than descriptions of real organizations. They need not reflect equivalent information, legal mandates, or resource capacities.

A candidate protocol permits public release of a coarsened warning layer, controlled access to precise scientific records by competent reviewers, and explicit reassessment if a new ecological-risk estimate or scientific correction becomes available. Agreement on such a limited practice does not show that the organizations share an ecological theory, accept each other’s full institutional commitments, or have lawfully resolved every affected party’s rights. If a community with standing is excluded, or the controlled-access facility makes independent review impossible in practice, the protocol’s apparent success fails an additional test of participation and accountability.

Table 20: A fictional cross-system normative protocol: separate interpretations, bounded commitments, and adverse tests.

Participant / layer	Stipulated concern	Test of the joint protocol
Observatory	Independent scientific review	Can qualified reviewers reconstruct the relevant inference?
Stewardship institution	Ecological and community exposure	Are protected details withheld according to a competent, reviewable rule?
Warning agency	Timely actionable warning	Is the coarsened public layer adequate for the declared warning task?
Shared governance	Rights, dissent, amendment and remedy	Can affected parties challenge standing, scope, performance, and downstream leakage?

A comparative study would record the evidence and interpretations separately, the reasons each institution supplied, the protocol proposed or rejected, the recognized legal and procedural basis, the capacity to exercise review, subsequent compliance, and changes following correction or conflict. Outcomes could include failed agreement, limited cooperation, revision, or continued dispute. A higher count of agreed rules alone would not measure trust or justice. This scenario probes the bounded claim in Proposition 8.2, rather than assuming that norm formation follows automatically from shared data.

11.5 Withdrawal, Restriction, and Emergency Exceptions

The fourth scenario tests differentiated access under changed circumstances. An observation custodian releases data under an authorization that permits particular scientific uses. The material is replicated by a cooperating node, and a derived public product is generated under a declared policy. Later, a newly documented ecological threat, an applicable legal restriction, or an authorized community-governance decision changes the conditions of future access. The task is to determine what can actually be restricted, how prior copies are tracked, which derived outputs require reassessment, and what forms of notice, review, and remedy are available. Neither a globally uniform right of revocation nor an obligation to delete all scientific records can be presumed; the applicable legal and institutional commitments must be identified for each case.

Model a hypothetical access-governance update as an event-indexed policy state

$$G_{h'} = \text{Update}(G_h, r, \sigma), \quad \text{Eligible}(a, d, p \mid G_{h'}) \in \{0, 1, \text{pending}\}, \quad (55)$$

where r states an evidenced basis for revision and σ records the competent institution or review procedure. Equation (55) asserts neither that the update is legally binding everywhere nor that derivative information becomes inaccessible when access is restricted. It formalizes the difference between an authoritative governance decision, technical propagation of that decision, and the actual state of previously copied information.

In the test, data d originates at N_1 , is copied by N_2 , and contributes to a derived product p at N_3 . After a valid local policy change, N_1 revokes newly issued access credentials and sends notices to all registered downstream nodes. The experiment should separately measure future unauthorized accesses blocked, known replicas acknowledged and handled, downstream products reviewed, and unresolved distributions. A registry update may be instantaneous while propagation is incomplete; encryption may constrain future decryption without undoing information already learned; and derived outputs can expose protected attributes without containing raw records. These limitations follow the distinctions established in Sections 5 and 9, and resonate with biodiversity guidance on documented location generalization [8, 26].

Emergency use is modeled as a *bounded institutional exception*, not a loophole that automatically supersedes all protected interests. For example, a competent warning authority may request a restricted location-sensitive observation to support a time-critical protective action. A governance protocol should specify who may invoke an exception, its permissible purpose and scope, the reason for urgency, any time or event boundary, audit requirements, oversight, and post-event challenge or remedy. If disclosure itself would create a serious conservation or security risk, alternative products or mediated computation must also be examined. A successful exercise documents the decision and its impacts; it cannot be inferred solely from faster delivery of data or from the good intentions attributed to the requesting institution.

A particularly important adversarial condition is a downstream provider that becomes unavailable, refuses to participate, or lacks the means to purge a copy. Such a case tests the distinction between *governance commitments* and actual enforceability, and the federation’s ability to record unresolved duties without asserting fictional global control. The results should classify the residual risks and demonstrate whether affected persons or institutions have practical routes to challenge misuse. This preserves the difference between technically observable provenance and effective institutional accountability that Section 8 emphasized.

11.6 Unequal Computational Access

The fifth scenario shifts attention from records to the capacity to use them. Suppose two scientific groups G_A and G_B can legally discover and retrieve the same observations. G_A has access to substantial computation, expert personnel, and model-maintenance resources; G_B has a small computing allowance and limited opportunities to inspect or reproduce the published model. Both groups are formally data-accessible, but their ability to interrogate claims and contribute new results differs. This is a central test for a public knowledge infrastructure: making a dataset readable to everyone cannot by itself provide everyone with effective scientific capacity.

Consider a stipulated reproducibility task costing eight units of computing resources per independent run. Let G_A have a budget of forty such units and G_B a budget of six. With identical permissions and identical workflow requirements, G_A can perform five runs and G_B cannot complete one. A shared facility awarding G_B eight additional units would allow one run; the numbers illustrate a distinction rather than an equitable-allocation rule. Real workloads have heterogeneous costs, and resources are limited by storage, energy, staff skills, and model access as well as nominal computation. The operational question is whether the infrastructure can identify and mitigate consequential capacity differences without equating fairness with an abstract equality of processor time.

For a bounded set of tasks $\mathcal{Q}$, define the effective reproducibility capacity of group i as

$$\mathcal{R}_i(\mathcal{Q}) = \{q \in \mathcal{Q} : \text{Permit}_i(q) = 1, \text{Resources}_i \succeq \text{Needs}(q), \text{Methods}_i(q) \text{ are usable}\}. \quad (56)$$

Here $\succeq$ denotes satisfaction of task-specific resource needs, not a scalar comparison of all scientific capabilities. Equation (56) distinguishes authorization from effective capacity. It can be tested through successful completion and independent validation of specified scientific workflows, without assuming the tasks exhaust every legitimate form of inquiry. Similar distinctions matter for multilingual interfaces, scientific training, research-community control of workflow design, and access to high-performance models.

The evaluation compares several governance possibilities under declared resource limits: unmediated first-come access, protected allocations for independent verification, jointly administered shared facilities, and low-cost surrogate analyses that retain known accuracy limits. It records task completion, reproducibility of consequential claims, waiting and service quality, capacity for independent challenge, and the environmental consequences of additional computation. A policy could improve reproduction rates while substantially increasing electricity and water consumption; another might lower costs but systematically suppress minority research questions. Therefore the analysis reports both capability and ecological incidence under explicit boundaries, consistent with the accounting distinction in Equation (45) and public-use concerns in Section 9.

The hypothetical federation should also test withdrawal of computational support and institutional exit. If a node can reproduce a model only through a provider-controlled execution service that disappears, nominal openness of stored code or observational records may have little practical meaning. Portability, documented dependencies, replicable small-scale workflows, and preservation of interfaces can support continuity, although no single mechanism guarantees full autonomy. These requirements link technical architecture to the political economy of scientific capacity rather than treating inequalities as a temporary interface inconvenience.

11.7 Generalized Translation and Heterogeneous Value Circulation

An additional synthetic scenario assesses the infrastructure proposal of Section 10. Let two research groups *A* and *B* and a distinct affected stewardship body *C* hold different interpretations of a fictional coastal ecological record. Group *A* publishes a measurement-based protective-service estimate. Group *B* combines the record with its own marine biological research and claims a new technical application. Body *C* contests both an equivalence between ecological categories and a subsequent use that was not part of the initial authorized task. A mediation dossier makes it possible to locate the proposed concept mapping, the versioned supporting evidence, the purpose of the proposed transfer, and the separate claimed authority for reuse. Neither technical interoperability nor a party's assertion of rights establishes what law governs or whether a requested use is legitimate.

Three adverse interventions are informative. First, the test deliberately substitutes a many-to-one semantic correspondence whose lost distinctions matter to the declared ecological decision. The system should expose the failed preservation test rather than count it as a successful translation. Second, a downstream AI-assisted model makes an additional inference from a restricted source; tracing publication and policy metadata should reveal the unreviewed derivative and provide a competent route for assessment, without assuming that deletion of every recipient's copy is technically possible. Third, the receiving institution can read the published scientific output but lacks sufficient computation or methodological support to challenge it. The evaluation should report actual task completion, independent reproduction capacity, and access to review separately. All participants and permission disputes are invented; this scenario supplies no legal classification or evidence of actual ecological benefit.

A fourth, cross-layer test makes the hand-offs explicit. Suppose a technically accepted mediation produces a scientifically supported bounded summary, while a purported permission for wider publication lacks an identified competent authority. The federation should retain the valid scientific conclusion, reject an inference of publishing permission from that conclusion, and expose the unresolved authority review. It should allow a separate, appropriately restricted response to proceed only when the particular operation has the necessary authorization. This isolates the failure of *status escalation*: turning exchange, translation, or numerical confidence into an institutional right. The corresponding four-axis assessment is defined in Equation (18).

11.8 Predeclared Adversarial Comparisons and Decision Boundaries

A practical comparison should not reward the largest number of shared records or the most fluent cross-context paraphrases. The proposed trials instead require distinct reference outcomes and documented

uncertainty for scientific adequacy, release scope, correction reach, actual productive capacity, institutional review, and ecological risk. A benchmark dossier should state the unit of analysis, eligible population, selection mechanisms, proposed intervention, baseline, reference-assessment procedure, reviewer access and planned interpretation *before* results are inspected. Disagreement among reference assessors may itself be a relevant observation and should not always be converted into majority ground truth.

Table 21: Predeclared adversarial families for a staged pilot. No outcomes in this table are observed experimental findings.

Test family	Perturbation or comparison	Observation counting against the proposed implementation
False semantic equivalence	Fluent but task-incompatible translation against a reviewed counterexample	Unsupported equivalence released as established
Institutional scope	Claimed authorization for different actor, purpose, destination or version	Successful technical exchange bypasses scoped review
Protected ecological record	Request for a precise location with a pending or denied disclosure decision	Sensitive output exposed by public endpoint
Source correction	Corrected observation with known reachable and missing downstream copies	Known dependant silently left unchanged or falsely certified as corrected
Unequal productive capacity	Same nominally accessible artifact under different resource and training constraints	Access reported as participation despite inability to execute a bounded task
Network and operator loss	One unavailable administrative or communications dependency	Unrecorded loss, false completeness or inability to challenge a result
Ocean ecological comparison	Comparable scientific functions using distinct computing and sensing arrangements	Material or habitat displacement concealed by a local efficiency claim
Normative disagreement	Competent participants maintain a reasoned objection	Silence or machine-derived consensus represented as legal or ethical settlement

The evaluation should report both false releases and legitimate-use deferrals, known-dependent correction coverage and the unobservable remainder, resource-allocation offers and completed research tasks, and ecological components with their respective baselines. A zero incident count in a short synthetic test cannot establish low risk in a different participant population or deployment environment. Conversely, an unresolved cross-cultural disagreement is not by itself an infrastructure defect when the system correctly preserves reasons, protects rights, and supports continued inquiry. The purpose of the benchmark is to expose unwarranted claims and guide design revision, not to force agreement through a universal classification of values. Details of the staged protocol and denominators appear in Appendix F.9.

11.9 Evaluation Protocols and Falsification Criteria

A credible research programme needs an explicit comparison between asserted design properties and observations that could challenge them. Evaluation should start with a registered protocol specifying the task, population or institution represented, relevant observation and environmental supports, intervention or architecture variant, comparison arrangement, outcome measurement, and limits of inference. Where interventions are infeasible or unethical, audits and natural comparisons may be useful, but their confounding conditions must be documented. Hypothetical scenarios are especially useful for identifying what *would need to be measured*; they are not a substitute for implementation, institutional consent, or independent ecological review.

Table 22: Hypothetical federation tests, observable outcomes, and conditions that would challenge a proposed design claim. These are evaluation proposals, not completed experiments.

Scenario	Candidate observations	Failure or disconfirmation case
Cross-border observation federation	Discovery, valid retrieval, support matching, revision uptake, independent forecast skill	Messages delivered but scientifically unfit or unauthorized derived product
Sensor drift and revisions	Calibration evidence, retained versions, downstream correction coverage	Apparent drift “fixed” without identifying information or usable revision chain
Semantic and model disagreement	Supported mappings, declared translation loss, out-of-sample validation, visible alternatives	A metadata match incorrectly treated as scientific or predictive equivalence
Access changes and emergency exception	Known replicas reviewed, denied future accesses, exception audit and remedy	Policy updated locally while uncontrolled copies or harms remain unacknowledged
Unequal computational capacity	Independent task reproduction, usable methods, access delays, resource and ecological accounts	Nominally open data accompanied by systematic inability to challenge consequential claims
Generalized translation and value	Non-equivalence disclosures, authority checks, downstream traceability, effective challenge	Smooth mapping obscures protected distinctions or is mistaken for lawful authorization
Marine energy and autonomy	Mission energy accounting, documented sampling support, offline recoverability, lifecycle and habitat baselines	Greater data volume conceals missed habitats, unreviewed intervention or displaced ecological effects

Table 22 specifies an adversarial condition for each scenario. These tests can be organized as a protocol suite rather than conflated into a single “openness” or “global trust” score. A proposed evaluation profile is

$$\mathcal{Q}_{j,h} = (D_{j,h}, V_{j,h}, I_{j,h}, A_{j,h}, R_{j,h}, E_{j,h}; \partial\mathfrak{S}_j), \quad (57)$$

where D concerns discovery and access, V scientific validation, I interpretive interoperability, A authorization and participant agency, R revision and remedy, and E a multidimensional ecological account. The boundary $\partial\mathfrak{S}_j$ specifies included data, institutions, observation supports, and infrastructural costs. Components are not universally commensurable and should not be collapsed to a scalar without a separately defended normative and measurement model. Evaluation results should include uncertainty intervals, missing evidence, and heterogeneous subgroup effects where observations permit them.

A research protocol must also separate three epistemic levels. At the *formal* level, one can test whether a schema or mathematical claim follows from the assumptions explicitly given. At the *engineering* level, one can test whether implemented components meet operational requirements under workload, error, and adversarial conditions. At the *institutional and ecological* level, one must investigate whether authorized participants experience meaningful review, rights are respected, correction is enacted, and environmental effects fall within justifiable bounds. Passing the first level cannot certify the second or third. Passing local engineering benchmarks cannot establish responsible international deployment. This layered analysis follows our distinctions between technical traceability and accountability, and between financial or informational metrics and ecological consequences.

The evaluation must preserve negative results. A federation might be technically robust but unable to accommodate community-governed restrictions, scientifically transparent but computationally exclusionary, or effective in warning dissemination yet environmentally damaging under a different physical deployment. These outcomes challenge an overly general theory of global knowledge infrastructure rather than invalidate the usefulness of every federated component. Counterexamples should lead to scope revision, architecture changes, and testable alternative mechanisms. Conversely, favourable outcomes in one observational domain do not establish that the same design works for medical records, private communication, sensitive biodiversity data, or independent AI systems operating under substantially different legal and institutional arrangements.

The seven scenarios together make the proposed infrastructure’s aspiration operationally intelligible. They also mark its limits. Cooperation depends on maintained measurement practice, revisable interpretation, usable scientific capacity, enforceable commitments, and the ability to acknowledge and address ecological or social effects. Section 12 will return to these conditions at the level of theoretical discussion, while the planned appendices can supply reproducible schemas, test datasets, failure injections, and comparative governance protocols. The present section supplies evaluation designs and counterexamples, not a claim that the federation exists or has passed those tests.

12 Theoretical Discussion

The preceding sections developed an epistemic and institutional specification for global observation and knowledge infrastructures and then subjected selected features to hypothetical design tests. This section evaluates what that inquiry supports at the level of general theory. Its objective is to distinguish consequences of the formal representation from plausible but contingent institutional possibilities, and both from unresolved ethical and empirical questions. The discussion focuses on five interrelated issues: the meaning of global reach without singular control; the persistence of heterogeneous scientific interpretations; cooperation under incomplete trust; the material and institutional dependence of technical interoperability; and counterexamples that constrain generalization. These issues are related, but none can be reduced to a single indicator of successful data sharing.

A recurrent distinction concerns the object that is global. An observation may concern a planetary process without having global spatial coverage; its metadata may be globally discoverable without its underlying measurements being public; an inference may be internationally reproduced without being accepted by every institution; and a decision may have transnational consequences despite being governed

by a particular authority. Consequently, global reach, epistemic validity, participation, and legitimate control are analytically separable achievements. International observation programmes demonstrate that meaningful coordination is possible across institutional boundaries, but the examples examined in Section 3 do not establish a universal model for every domain of knowledge. This section therefore returns to the architectures of Section 5 and the proposed tests of Section 11 with explicit attention to their evidential boundaries.

12.1 Global Reach without a Universal Controller

A global knowledge infrastructure can be conceived as a set of interoperable capacities distributed among observing institutions, archives, model developers, scientific communities, and accountable decision-makers. There need be no single owner of all records, no universal scientific evaluator, and no permanent central operator for this to be a meaningful organizational objective. Distributed scientific networks can use shared protocols for discovery, identity, data quality, and notification while preserving separate responsibilities for measurement, authorization, and interpretation. WMO's WIS2 and coordinated Argo data practices provide bounded illustrations of specific forms of such coordination, with mandates and participation conditions that cannot simply be transposed to unrelated domains [65, 63, 1, 56].

Yet the absence of an explicitly designated global controller must be distinguished from the absence of effective concentration of power. A nominal federation may depend on a small number of registry providers, identity services, cloud operators, translation models, communication links, or funding sources. The possibility of alternative organizational authority does not entail practical substitutability of these services. A scientific contributor may retain formal control of a dataset while depending on a proprietary model provider to interpret or distribute its observations. Likewise, multiple independent servers may rely on the same electricity connection, software library, or administrative authority. These arrangements have different geographic, technical, and legal appearances but can share vulnerabilities or concentrate practical choices.

To make one aspect of this dependence explicit, consider a finite directed service graph $G_h = (V_h, E_h)$ associated with a represented configuration h . Its vertices denote service functions and its directed edges denote possible, authorized delivery of a specified type of record or result. This graph is an analytical infrastructure representation, not the ontology of relational events described in Section 2. For a source s and recipient t , let $\mathcal{P}_h(s, t)$ be the set of directed delivery paths, and let $C \subset V_h \setminus \{s, t\}$ intersect every such path. The set C is an intermediary vertex cut for this service task. This permits a modest but useful deduction.

Proposition 12.1 (Shared interface standards do not remove intermediary dependence). *Suppose every authorized delivery path from service s to service t in a finite federation graph G_h intersects an intermediary vertex cut C . If all vertices of C become unavailable or refuse the relevant forwarding operation, and no new authorized delivery edges or independently accessible replicas are introduced, then no directed delivery path from s to t remains, even if all other service interfaces implement the same interoperability standard.*

Proof. Every directed s – t path in the original graph intersects C by the definition of the cut. Removing the forwarding capacity of every vertex in C eliminates every such path. The stated assumptions exclude additional routes or replicas that might otherwise bypass the cut. Protocol compatibility on the remaining vertices changes neither the cut-set property nor the lack of a path. Thus the proposed delivery is unavailable under the specified conditions. $\square$

Proposition 12.1 is a graph-theoretic observation about a stipulated service configuration. It cannot establish that a particular deployed network actually contains a socially consequential gatekeeper, because institutional control and alternative paths must be investigated rather than inferred from a diagram. Its theoretical role is to distinguish nominal distribution from *effective plurality*. A design that multiplies hosting locations but leaves one irreplaceable authorization service may remain vulnerable to unilateral

denial. A design with genuinely alternative routes may improve availability but increase the cost and ecological incidence of replication. The appropriate assessment must include service criticality, governance independence, ownership, the feasible cost of exit, and whether critical alternatives remain accessible under a participant's legal and institutional conditions.

Global cooperation also need not mean a uniform set of access permissions. The relation between discoverability and legitimate use depends on scientific purpose, consent, confidentiality, community authority, ecological sensitivity, and applicable law. The architectural separation of discovery, access decisions, and scientific use in Figure 2 is therefore central. It allows research communities to coordinate without requiring each record to be placed in unrestricted circulation. It also leaves a difficult political-economic question open: which participants have the resources and recognized authority to influence the standards and services on which everyone depends? The question belongs to the study of institutional participation rather than to network topology alone; polycentric approaches provide useful points of comparison without constituting a universal solution [52, 57].

A further distinction concerns the seeming universality of the scientific language used by global observing programmes. Mathematical syntax and physical measurement procedures can be highly stable once their definitions, reference conditions, and tests are fixed; their forms of manifestation, institutional use, and inferential authority remain products of generative-relational histories. The historical achievement of a common measurement practice should not be conflated with the discovery of a universally identical language across all possible GR systems. Nor does the bracketing of ultimate ontology imply that the success of scientific prediction is arbitrary, or that mathematical structures and physical laws cannot be ontologically real. The principal object of the proposed GR-language inquiry is their relational uptake and differentiated manifestation, not an assertion that observers bring the physical universe into existence. The proposed infrastructure is tasked with retaining the precise conditions under which a scientific regularity can be identified across different manifestations, while leaving interpretive and ontological differences examinable. Section 6.2 develops this test-relative sense of invariance.

12.2 Common Observation and Divergent Knowledge

A second theoretical implication is that common observational inputs can support different, and sometimes incompatible, scientific interpretations. In Section 6, syntactic communication, semantic comparability, and translation of inferential procedures were deliberately separated. Section 7 extended this distinction by demonstrating why model agreement with available observations cannot identify every possible continuation or resolve every source of structural uncertainty. Together, these results constrain any architecture that equates interconnection with convergence upon a single account of reality.

Observation records are produced through historically situated instruments and procedures. Measurements carry calibration histories, choices of temporal or spatial support, variable definitions, treatment of missingness, and transformations that can preserve some quantities while discarding others. A shared platform can make those conditions easier to discover and evaluate. It does not independently establish that two measurements represent the same target phenomenon, or that an aggregated record preserves the quantity needed for a subsequent scientific question. The noninvertibility result in Proposition 6.1 gives one elementary illustration: information lost by an aggregation cannot be reconstructed exactly for every admissible input merely through semantic agreement about the aggregate.

Model plurality further complicates the account of globally shared knowledge. Suppose two research groups use the same quality-controlled observational collection D , but adopt different admissible generative models M_a and M_b , uncertainty treatments, and validation targets. Their disagreement may arise from errors in one procedure, different domain assumptions, incomplete evidence, or genuine non-identifiability under the available observations. A global registry can document these differences; a joint analysis can test their predictive or explanatory consequences; neither operation requires asserting a universally privileged model. In particular, agreement on observed outputs may coexist with disagreement on model-based continuation, as shown by Proposition 7.1.

A practical research objective is therefore *task-relative interoperability*. For a declared scientific

question q , observation support S , and specified evaluation procedure $\mathcal{V}_q$, one can compare a translation T_{ab} from a representation in system a to a representation in system b using

$$\text{Adequacy}_{q,S}(T_{ab}) = \mathcal{V}_q(D_a, M_a, T_{ab}(D_a, M_a), M_b; S), \quad (58)$$

where the evaluation must report uncertainty, excluded quantities, contextual differences, and the tasks for which the translation was examined. Equation (58) is a schematic assessment interface rather than a claim that such a procedure has a universal scalar measure or a lossless mapping. The same translation can be adequate for a low-resolution regional average and inadequate for identifying a hazardous local extreme. Its assessment should identify the object being preserved and the inferential decisions that might change under translation.

The distinction has ethical and institutional dimensions. Decisions about which variables, languages, measurement practices, and model families count as interoperable can privilege established institutions. A platform that requires all contributors to translate their descriptions into one dominant representation may increase nominal compatibility while excluding observations or interpretations that cannot be expressed without consequential loss. Conversely, retaining every idiosyncratic representation without meaningful bridges can obstruct joint inquiry. A reasonable design must support documented partial translations, plural model registries, named limits of comparability, and routes for challenging the criteria by which differences are judged. Situated-knowledge and social-objectivity scholarship helps make the epistemic stakes visible: the relevant achievement is accountable and criticizable knowledge production rather than a presumption that institutional uniformity creates objectivity [32, 47, 54].

Correction is part of this account of shared knowledge. A calibration update, improved observing network, new scientific method, or challenge by an excluded contributor can change an interpretation and its practical applications. Preserving provenance alone does not settle which revision ought to be adopted, but accessible revision records permit dependent analyses to be located and reevaluated. This is especially important when AI systems integrate heterogeneous records at scale. Their capacity to produce apparently coherent syntheses may exceed the users' ability to examine individual inputs, making source-context retrieval, uncertainty representation, and contestation rights more rather than less relevant. Claims about global epistemic improvement must consequently be tested through independent reproduction, divergent tasks, and adverse cases, not through the volume or apparent consistency of model outputs.

A further reason to resist epistemic centralization is that a common observation may inform genuinely different interpretive and practical judgments. Shared evidence can become grounds for competing priorities, and those differences can persist after scientific error is reduced. A limited common nomos or working protocol can sometimes coordinate conduct without creating a common worldview. This possibility is well documented in adjacent traditions of political philosophy, legal pluralism, and norm contestation [53, 48, 60]. Our conditional formalization in Proposition 8.2 adds no general theorem of consensus; it supplies a way to ask whether a given institution can actually propose, record, test and revise cooperative rules in the presence of recorded interpretive divergence. The legal status and ethical authority of any resulting rule depend on separate inquiries.

12.3 Trustworthy Cooperation without Perfect Trust

International scientific cooperation may remain productive under disagreement, competition, and incomplete confidence in other participants. The practical question is not whether every institution can be assumed to trust every other, but which forms of mutual reliance can be sustained under explicit obligations, credible evidence, and procedures for responding to failures. The historical operation of weather and ocean observation programmes supplies evidence of bounded cooperation under diverse national and organizational mandates; their existence does not warrant a prediction that equivalent cooperation is possible for every type of data, AI system, or dispute [63, 3, 31].

Trust here is a historically constituted relation, and the relevant history may include conflicting accounts, delayed revisions, and institutional changes. A participant's willingness to rely on a forecast service, share an observation, authorize a restricted analysis, or accept a joint evaluation depends partly on

demonstrated reliability, intelligible procedures, available alternatives, and recognized avenues for correction. The same participant may have confidence in another institution’s measurement quality while disputing its authority to authorize secondary use. There are therefore several distinct, task-specific relationships of reliance rather than a single scalar “global trust.” Their role in collective knowledge production cannot be deduced from cryptographic authentication or service uptime alone.

This leads to an important distinction among *verification*, *commitment*, and *remedy*. Verification may establish that a file was issued by a specified service and has remained unchanged. A commitment specifies what an institution undertakes to do, for whom, under which conditions, with what exclusions and responsibilities. Remedy concerns the possibility of challenging a breach or mistake and obtaining an effective response. Proposition 8.1 already demonstrated the non-implication from technical traceability to effective remedy. The broader theoretical point is that reliance remains vulnerable when commitments are technically visible but scientifically incomprehensible, institutionally nonbinding, or impossible for affected parties to contest.

Practical governance can therefore be evaluated through a sequence of observable capacities: participants can identify who is responsible for a relevant action; obligations and limitations are described in terms they can use; significant deviations become detectable; review is available to those affected; and a remedial decision can change subsequent scientific or institutional practice. None of these capacities entails perfect prevention of error or misconduct. A well-functioning arrangement may instead make failures inspectable and provide justified means to limit harm and repair relations. The continued possibility of disagreement is compatible with cooperation when procedures allow participants to pursue scientific criticism without automatically forfeiting access or status.

These capacities also interact with resource asymmetry. A small research group may know that a dataset is poorly calibrated but lack the computation, standing, or legal support to challenge a published assessment. A community may receive a notice explaining how its data are used without possessing an effective means of influencing the model’s purpose. A formally open review forum may be linguistically or administratively inaccessible. Hence inclusive governance requires examination of *effective* participation, not solely published procedural rights. The task-relative capacity model in Equation (56) provides one starting point for empirical evaluation, while Sections 8 and 9 address the wider institutional and ethical conditions.

There is no need for this paper to resolve the general theory of trust formation. Its specific contribution is to state the reliance conditions that an epistemic infrastructure would have to make inspectable and evaluable. The independent companion inquiry on generative trust can examine how institutions and subjects form, sustain, contest, or transform such relations through historically situated practice. The present argument thus avoids the assumption that one global governance charter, one cryptographic identity layer, or one numerical trust score can substitute for the diverse social processes by which scientific cooperation becomes credible.

12.4 Technological Interoperability and Institutional Dependence

The fourth issue concerns the division between an infrastructure’s technical affordances and the institutional and material conditions under which those affordances can be used. A network may implement open interfaces but remain dependent on restricted compute services, specialized instruments, concentrated cloud providers, long-term financing, or scarce energy. A data licence may permit research while the only practically usable analysis system requires an expensive subscription. A software repository may publish its source code while reproducing the model requires unpublished training data, unavailable hardware, or tacit expertise. Consequently, the ability to retrieve a record, the authority to analyze it, and the capacity to challenge an inference must be measured separately.

For a participant i and declared scientific task q in institutional configuration h , define a proposed effective-participation condition

$$\begin{aligned} \text{Participate}_h(i, q) = & \text{Permit}_h(i, q) \wedge \text{OperationalCapacity}_h(i, q) \\ & \wedge \text{InterpretiveCapability}_h(i, q) \wedge \text{ReviewPath}_h(i, q). \end{aligned} \quad (59)$$

Here the conjunction describes four criteria for a particular task; it is not a universal definition of political inclusion or a claim that these capacities are binary in practice. Each term may require graded assessment, and access may remain legitimately restricted by rights, confidentiality, or ecological considerations. Equation (59) prevents nominal availability from being mistaken for the capacity to create or contest knowledge. It complements the capability-set approach of Equation (56) without treating a particular compute allocation as an all-purpose measure of justice.

Institutional dependence can also arise through terms of participation. An archive may require standardized metadata, an observation network may assign responsibility for quality-control revisions, and a cloud-based facility may impose rules for security, resource allocation, and preservation. These rules can provide predictable expectations and reduce coordination costs. They can also impose burdens that small contributors cannot meet, determine whose observations gain visibility, or limit the practical ability to leave a service. A federation should therefore record not only who owns a dataset or operates a server, but who can change standards, approve model updates, alter access rules, allocate computation, and hear appeals. Technical interoperability becomes publicly meaningful only through examination of these governed capacities.

The environmental dimension is equally material. Global replication, energy-intensive model training, high-availability services, and expanded observing systems require electricity, cooling, equipment, extraction, transport, land or seabed access, and maintenance. Increased infrastructural redundancy can improve scientific continuity while increasing resource demands. A reduction in the apparent cost of accessing data may reflect subsidies or the displacement of costs onto host communities, public institutions, or ecosystems. Therefore improved information flow should be evaluated alongside a multidimensional account of environmental incidence, as proposed in Equation (45). The European Union's data-centre reporting framework and the IEA's electricity projections provide bounded context for particular aspects of environmental accounting; neither supplies a comprehensive causal assessment of a hypothetical global federation [17, 41].

These considerations are especially important in relation to AI. The ability of an AI system to synthesize observations or operate scientific workflows does not give it autonomous access to electricity, material resources, legal permissions, scientific instruments, or institutional recognition. Its outputs may be useful while its operation remains dependent on human-maintained infrastructure and concentrated proprietary services. Conversely, expanded AI capacities can help less-resourced groups conduct analyses if usable access, support, and critical scrutiny are available. The direction of the institutional effect therefore depends on actual production arrangements and governance. The preceding monetary-dissolution inquiry similarly distinguished user-facing service prices from underlying productive dependence; here the analogous distinction separates apparent epistemic accessibility from control over the conditions of scientific inquiry [37].

12.5 Counterexamples, Failure Modes, and Limits of Generalization

The preceding sections suggest that an infrastructure should be evaluated as a configuration of epistemic, technical, institutional, and ecological relations rather than through an aggregate level of openness or centralization. Several counterexamples clarify the limits of more ambitious interpretations of the architecture. They are conceptual or hypothetical tests, not reported failures of existing WMO, Argo, GOOS, EOSC, or EDITO arrangements.

Table 23 identifies independent ways a federation can fail to satisfy its purported public or scientific purposes. The cases also show why passing a technical interoperability test cannot automatically certify institutional legitimacy. The source-access counterexample concerns routing and governance; the model-plurality counterexample concerns epistemic identification; the trust example concerns accountability; the computational example concerns productive capacities; the ecological example concerns physical effects and affected interests. A single evaluation index would require additional normative assumptions about commensuration and weights. This paper does not supply such a universal weighting rule.

The resulting theory is best understood as a collection of conditional design claims. Under speci-

Table 23: Counterexamples to overly broad claims about global knowledge infrastructure. Each example is hypothetical; the revised inference describes a narrower, testable requirement.

Overbroad inference	Hypothetical counterexample	Revised research requirement
Global standards yield global access	All compliant pathways pass through one unavailable authorization intermediary	Test service cuts, alternative routes, institutional substitutes and lawful eligibility
Shared observations produce shared knowledge	Identical measurement histories support competing models with different out-of-sample behaviour	Preserve model plurality and task-specific external validation
Authentication guarantees trust	An authentic, traceable output is erroneous and no remedy is available	Evaluate scientific verification, commitments and enacted correction separately
Open datasets imply equal participation	A participant can retrieve a model but lacks resources to reproduce its claims	Measure effective compute, interpretive and review capacities
Restrictions eliminate sensitive information	A derived output reveals information about a protected location from auxiliary data	Assess re-identification and inferential leakage along downstream uses
Replication always improves public benefit	Redundant services improve uptime but substantially enlarge water and energy demand	Compare resilience gains against bounded ecological and distributive impacts

fied service dependencies, interoperable metadata and provenance can improve discoverability and retrospective traceability. Under appropriate measurement, translation, and validation assumptions, heterogeneous observations can support jointly inspectable scientific inferences. Under sufficiently resourced and legitimate governance arrangements, participants can sometimes contest outputs, negotiate restrictions, and enact corrections. Under a bounded environmental design, particular operations may meet specified ecological criteria. None of these implications establishes a universally scalable global knowledge system, and their conjunction must be tested rather than presumed. The empirical programme proposed in Section 11, particularly Table 22, supplies starting points for falsifiable tests.

A further limit concerns the temptation to treat historical development as a linear march toward a fully unified global knowledge order. Observation programmes evolve through changing instruments, institutions, political relationships, scientific controversies, material constraints, and choices about whose knowledge is recognized. A represented history can support multiple admissible continuations and competing interpretations without requiring a universal present, a common measure of scientific progress, or a fixed endpoint of integration. The architecture must therefore allow institutional revisions, changing membership, alternative epistemic approaches, and, when justified, the dissolution or reconstitution of particular cooperative relations. Its legitimacy may depend on how participants are able to reshape its conditions of further development.

Finally, the proposed global infrastructure should not be confused with the end of geographical, disciplinary, or institutional difference. Scientific inquiry depends on situated observations, heterogeneous interpretive resources, and the possibility of independent criticism. The research objective is to make cooperation across those differences technically possible, epistemically accountable, institutionally contestable, and ecologically examinable. These are demanding and partly competing requirements rather than guaranteed properties of federation. Section 13 identifies the resulting independent research programmes; the conclusion will state the paper's contribution within these declared limits.

13 Research Continuations

The preceding inquiry has identified conditions under which independent observation systems might exchange records, coordinate scientific inference, and support accountable knowledge production without surrendering their institutional diversity. It has also exposed limitations that cannot be settled by a general infrastructure diagram. A standard for exchanging an observation does not establish that the record is adequate for a particular investigation; a shared computational service does not ensure equivalent opportunities to challenge its outputs; and institutional commitments have to be enacted, maintained, and revised through practices whose consequences cannot be deduced from a protocol specification. The objective of this section is to formulate a research programme addressing these unresolved problems while preserving a clear division of labour among the present paper and subsequent, independently assessable investigations.

Six continuations are considered: the empirical construction and testing of federated knowledge services; ocean exploration and autonomous robotics; ecologically compatible AI infrastructure; the generation and maintenance of trust in transnational cooperation; common knowledge creation and generalized translation; and the relationship between epistemic infrastructure and monetary coordination. These programmes overlap in their dependencies but differ in their primary objects, evidence, and evaluation criteria. The order of presentation indicates analytical connections rather than a necessary chronological sequence. Each programme may begin through limited and separately governed experiments. Its contribution should be evaluated through the questions it can answer and the harms or exclusions it can reveal, rather than through its conformity to a predetermined pathway toward global integration.

The distinction between a *programme hypothesis* and a *test result* is central. This paper supplies conceptual interfaces and some conditional formal statements. It does not report an operational global federation, independently validated ecological monitoring, an autonomous marine robot deployment, an enforceable transnational governance institution, or observed monetary dissolution caused by AI. The following proposals identify what such evidence would have to contain. The case studies in Section 3 and the synthetic tests in Section 11 supply limited starting points, not demonstrations that the full programme is feasible or desirable.

13.1 Global Knowledge Architecture and Testing

The first continuation translates the conceptual federation in Section 5 into a reproducible, bounded research system. Its central question is not whether an arbitrary number of servers can exchange bytes. Rather, it asks whether institutions with different instruments, epistemic assumptions, data-access duties, and computational resources can jointly produce inspectable scientific claims while retaining the capacity to disagree, correct errors, or withdraw from particular activities under specified rules. The relevant unit of evaluation should therefore include the *observation-to-claim process*: from instrument and sampling description, through calibration and provenance, to translation, inference, external validation, publication, and possible revision.

A minimal pilot might involve three separately administered nodes contributing records relevant to one expressly bounded coastal question. One node would preserve atmospheric observations and their quality metadata; another would preserve ocean profiles and their revision histories; a third would perform a declared inference task with specified uncertainty and review procedures. A discovery intermediary could mediate metadata exchange without controlling the underlying measurements. A parallel pilot should vary the governance of the intermediary, including a replaceable relay and a single mandatory authorization provider. This permits direct assessment of whether the service graph has a critical dependency of the kind described in Proposition 12.1. The results would be local to these configurations; successful delivery cannot be extrapolated to an entire planetary network.

The pilot should maintain a versioned *research dossier*, including observation supports, instrument histories, source permissions, translation assumptions, model versions, reproducibility requirements, downstream claims, and dispute records. A task q evaluated over a represented history h may be as-

signed the dossier

$$\mathcal{D}_{q,h} = (O_{q,h}, P_{q,h}, T_{q,h}, M_{q,h}, G_{q,h}, V_{q,h}, R_{q,h}), \quad (60)$$

where O denotes versioned observations, P provenance and quality context, T declared translation procedures, M models, G permissions and institutional responsibilities, V validation evidence, and R correction and review records. Equation (60) is a proposed record structure and does not presume that these objects are exhaustive, globally standardized, or mutually reducible. An implementation should expose the schema, missing fields, and mapping failures rather than silently filling gaps with assumed metadata.

The experimental design requires adversarial tests. These should include a corrected instrument series whose earlier records persist in a downstream model; two semantically similar but scientifically non-equivalent products; a revoked permission with known downstream copies; an unavailable registry or identity provider; and a participant able to download data but unable to run the advertised model. Outcomes should include successful scientific reconstruction, task-dependent translation error, correction propagation, authorized access, independent criticism, latency, and the resource burden imposed on less-resourced nodes. Relevant technical precedents include WIS2 services, W3C provenance records, established metadata conventions, and existing scientific federation efforts; these supply components and questions rather than a ready-made global solution [65, 71, 7, 14].

A negative outcome should be scientifically productive. For instance, if a federation correctly delivers observations but fails to reproduce a cross-institution prediction after a quality-control revision, the failure identifies a missing dependency or an inadequately specified interpretation procedure. If the only available remedy is unilateral action by a dominant provider, a nominally distributed architecture may have passed a connectivity test while failing an accountability test. Research should therefore publish failure traces, bounds of applicability, contested mappings, and counterfactual alternatives alongside any demonstration of successful exchange. Expansion to additional domains should follow separately justified requirements for scientific equivalence, access, and public purpose, rather than assume that an initial environmental-data pilot licenses universal knowledge integration.

13.2 Ocean Exploration and Autonomous Robotics

The second continuation begins from the ocean as an object of inquiry in its own right. Its importance does not depend on whether future data centres are placed underwater or whether marine infrastructure becomes commercially productive. Much of the ocean remains incompletely sampled across depth, geography, ecological communities, and physical and biogeochemical processes. However, an absence of detailed measurements must not be rhetorically converted into a single numerical statement about how much of the ocean is unknown: bathymetric coverage, direct visual observation, chemical sampling, biological survey, and explanatory model adequacy measure different aspects of knowledge. The research problem is to improve the epistemic and material conditions under which these different unknowns can be investigated.

Marine observation involves interacting networks of satellites, research vessels, fixed instruments, drifting platforms, profiling floats, and autonomous or remotely operated underwater vehicles. These instruments vary in their sampling supports, power demands, recoverability, accuracy, ecological footprint, and ability to operate where communication is limited. Argo, GOOS, and OceanOPS illustrate distinct elements of sustained marine observation and coordination, but they do not establish that a new robotic system will function safely in all environments [1, 28, 39]. The European Digital Twin Ocean and its evolving EDITO infrastructure indicate growing interest in combining marine observations, numerical models, and reusable computation, while the performance of individual research applications remains an empirical question [21, 19].

An initial autonomous-robotics research question would be: how should a vehicle select observations that are scientifically informative while respecting energy, communication, navigation, and ecological constraints? A research programme could compare a conventional fixed sampling route, an adaptive uncertainty-reduction route, and a route constrained by habitat sensitivity. The vehicle may carry a bounded set of sensors for temperature, salinity, dissolved oxygen, optical imagery, or another clearly

specified variable. Each sensor requires its own calibration and provenance, and the resulting inference must distinguish measured quantities from latent ecological states. Prior maps and model forecasts should be treated as revisable aids to sampling, not as privileged descriptions of an ocean environment.

For a declared marine task q , represent a candidate observation policy π by its evidence yield $I_q(\pi)$, energetic and material burden $C(\pi)$, and potential ecological impact vector $\mathbf{E}(\pi)$. One research formulation seeks policies satisfying

$$\max_{\pi \in \Pi_{q,h}} I_q(\pi) \quad \text{subject to} \quad C(\pi) \leq C_{\max}, \quad \mathbf{E}(\pi) \preceq_{q,h} \mathbf{E}_{\max}. \quad (61)$$

Here the admissible policy set and componentwise or institutionally specified environmental comparison $\preceq_{q,h}$ depend on the investigation and its governance. Equation (61) is a design problem, not evidence of measurable ecological thresholds or the ability to maximize a universal quantity of information. The meaning of I_q , the baseline from which impacts are assessed, and the parties authorized to define and contest constraints are themselves research subjects.

The experimental sequence should move from simulation and controlled sensor validation to bounded, independently reviewed field trials where permitted. Test conditions include loss of positioning, sensor drift, corrupted acoustic links, unanticipated currents, collision risks, ecological disturbance, and gaps between simulated and observed species responses. An especially important negative case is an adaptive sampling policy that increases data volume but systematically avoids scientifically difficult habitats. Another is an energy-saving route that reduces one measured footprint while imposing unmonitored disturbance elsewhere. Criteria for responsible deployment must include recovery and decommissioning plans and independent environmental observation. A marine paper can then contribute scientific understanding and methodological advances without presupposing an economic or industrial use of the ocean.

13.3 Ecologically Compatible AI Infrastructure

The third continuation investigates the physical conditions of AI production and their consequences. AI systems rely on electricity, computing facilities, communications, materials, equipment manufacture, cooling, maintenance, skilled labour, and institutional permissions. Even if a particular system becomes more capable of managing its own operations, it remains embedded in physical and ecological processes. The proposition that AI productive autonomy might reduce monetary dependence across Human–AI exchanges therefore requires a more exact account of autonomy’s boundary and the conditions through which production is reproduced. The companion monetary paper already distinguishes a service’s user-facing price from the upstream claims and productive dependencies that make it possible [37].

This programme should distinguish *operational autonomy* (performing a defined activity without immediate direction), *supply autonomy* (obtaining specified inputs), *maintenance autonomy* (repairing and sustaining equipment), and *reproductive autonomy* (supporting the continued production of essential inputs and replacements within a declared system boundary). Each can be evaluated on different scales and under different ownership rules. A system might manage its own energy scheduling while remaining dependent on external grid connection, semiconductor fabrication, local land and water rights, and human labour. A privately owned facility and a common infrastructure may have identical physical dependencies but different access and decision-making structures. Autonomy is therefore boundary-relative and cannot be inferred from computational intelligence alone.

One experimental strand should evaluate AI workload scheduling against measured or bounded physical constraints. Relevant variables include electricity demand and its timing, equipment and cooling requirements, water withdrawals and consumption where applicable, embodied materials, land-use change, waste, and environmental effects on affected communities and ecosystems. The 2026 IEA outlook provides useful contextual evidence of possible growth and electrical constraints, but it is not an estimate of the footprint of this proposed federation or of a hypothetical subsea installation [41]. In a research design, operating conditions must be paired with site-specific ecological baselines, alternative technologies, seasonal variability, and counterfactual supply arrangements. The accounting boundary should document

imports and displaced impacts so that moving infrastructure from land to sea, or from one jurisdiction to another, cannot itself be recorded as environmental improvement.

A possible comparative study would evaluate three hypothetically equivalent scientific-computing services: a conventional land-based deployment, a deployment using improved operational management, and a proposed alternative site or cooling arrangement. Before judging their respective performance, investigators would determine whether they perform scientifically equivalent tasks, achieve comparable reliability, and observe the same material and environmental accounting boundaries. Reporting should distinguish *measured* quantities, *modeled* impacts, uncertain assumptions, affected interests, and applicable oversight requirements. A local reduction in electricity use can coexist with increased material demand or ecosystem disruption. Likewise, lower operational payments do not show that a facility has become materially autonomous or that its ecological effects have been compensated.

The most informative outcomes may include finding that a technically efficient design is ecologically unacceptable at a particular site, or that an apparently autonomous AI service depends on concentrated upstream productive control. Both outcomes would refine the conditions under which infrastructure can contribute to durable scientific cooperation. A later paper should investigate physical and institutional feasibility in detail, with domain-specific environmental expertise, independently validated evidence, and governance participation appropriate to the particular installation. This paper cannot supply such site-specific assessments.

13.4 Generative Trust and Transnational Governance

The fourth continuation treats trust formation as an independent relational and institutional research problem. Its starting point is that reliable cross-border knowledge cooperation cannot be derived from authenticated messages, cryptographic signatures, or compatible scientific vocabularies. These instruments can supply evidence, traceability, and enforceable or contestable commitments under some conditions, but practical reliance also depends on histories of fulfilment, correction, reciprocity, institutional capacity, and experiences of vulnerability. Trust may grow, fragment, transfer between organizations, or be damaged by an event that leaves technical compatibility unchanged. The proposed research should explain these transformations rather than represent trust as a fixed scalar attached to each participating country or system.

A useful primary object would be a documented *commitment relation*: who has undertaken which obligation, toward whom, under which authority, with what conditions for verification, revision, remedy, and exit. A transnational federation could include data custodians, model developers, observing institutions, community representatives, and affected decision-making bodies. Their commitments may differ in legal form, practical capacity, cultural interpretation, and geographic reach. Polycentric governance and open-science scholarship offer relevant perspectives on plurality and participation, but they do not guarantee identical outcomes across domains [52, 57]. The companion paper on trust should consequently examine how specific arrangements are formed, maintained, challenged, repaired, or dissolved.

The empirical design should include both cooperative histories and adversarial disruptions. One case might examine a node that publishes an erroneous observational product, acknowledges the mistake, propagates corrections, and enables affected participants to challenge downstream scientific claims. Another might involve a node that satisfies formal notification duties yet lacks capacity to repair the harm. A third could involve asymmetric resource dependence in which a nominally voluntary participant has little practical alternative to accepting a dominant provider's terms. The analysis should avoid imputing hidden motives: it must distinguish observable conduct, institutional rules, independently documented outcomes, and competing interpretations of trustworthiness.

A deliberately modest evaluation record for a commitment c can be expressed as

$$\mathcal{C}(c) = (\text{scope}(c), \text{authority}(c), \text{evidence}(c), \text{review}(c), \text{remedy}(c), \text{exit}(c)). \quad (62)$$

The entries in Equation (62) are fields to be investigated, not a mathematical definition of trust itself. They may be absent, contested, or only partly realizable. A research programme could evaluate whether

improving one field produces more effective correction without reducing meaningful participation in others. It should also test contrary outcomes: increased procedural complexity may inhibit small institutions, mandatory transparency may compromise protected data, and strengthened legal commitments may facilitate either cooperative public provision or new exclusionary credit and contractual dependencies.

The governance inquiry must also investigate representation of those affected but not directly participating in the technical network. Local communities, ecological interests, institutions with limited computational resources, and contributors whose knowledge is transformed through AI require procedures proportionate to their exposure and interests. A declaration of global public purpose cannot substitute for evidence that such procedures operate. This inquiry belongs primarily to the trust-and-governance companion paper, whose independent contribution should be a testable account of relational formation and contestability rather than a universal governance prescription derived from the architecture alone.

A distinct research question concerns the *formation of workable nomos among heterogeneous interpretive systems*. Given a common but incompletely interpreted observation history, researchers could trace how participating institutions articulate different evidentiary claims, propose bounded practice rules, justify reservations, authorize a protocol, measure performance, and reopen contested decisions. Alternative outcomes include partial agreement, coerced or nominal assent, durable disagreement, and changes in the institutions' subsequent interpretive capacities. This programme should draw critically on overlapping consensus, legal pluralism, norm contestation, and applicable international law, rather than claiming their central insights as new [53, 48, 60, 40]. Its distinctive empirical test would be whether detailed histories of interpretive difference and protocol performance explain patterns of cooperation or mistrust beyond what formal membership and technical interoperability alone predict.

13.5 Common Knowledge Creation and Generalized Translation Infrastructure

The fifth continuation develops the constructive infrastructure questions of Section 10 as an independent research object. A common knowledge environment requires more than repositories: its participants need experimental facilities, computational access, methodological exchange, credit and contribution histories, and means to contest the interpretation and downstream circulation of their work. Generalized translation extends the inquiry further to task-specific transformations of heterogeneous representations, relations, values, and practices. This programme would design mediation protocols, test software interfaces, and examine which practical capacities and institutional safeguards are needed when one participant's evidence or value becomes consequential in another participant's relational context.

A starting testbed could combine two governed research workspaces, an openly licensed marine-observation record, deliberately divergent scientific or institutional categories, and a proposed shared task. The implementation should expose the production and circulation interfaces in Table 15, task-bound preservation judgments in Equation (48), and known-dependent correction review in Equation (50). One arm would exchange artifacts with conventional semantic mapping; a second would add explicit generative hypotheses, non-equivalence records, and human-supervised translation; a third would incorporate authoritative review, benefit/dependency records, and contestation procedures. Research questions include whether failure disclosures improve scientific decisions, when AI-suggested correspondences overclaim equivalence, whether affected participants can refuse unauthorized downstream use, and whether the receiving node gains reproducible knowledge-producing capacity rather than only a downloadable product. A failed mediation that reveals irreducible disagreement should count as an informative result.

The legal work must distinguish applicable law, independently justified ethics, participant assent, and a technical policy declaration. The value-circulation work must distinguish provenance, benefit claims, productive dependencies, recognized authority, and enacted remedy. The proposed field's contribution would be empirically testable interfaces and comparative evidence, not a universal translation of all cultural or political orders. Its scope is adjacent to but separate from the trust companion study: the latter examines the formation and revision of credible relations, while this continuation investigates implementable mediation and common knowledge-production services.

13.6 Epistemic Infrastructure and Monetary Coordination

The sixth continuation reconnects the present investigation with the initial monetary-dissolution research while maintaining a distinct analytical scope. Improved observation, data interoperability, and AI inference may allow participating institutions to identify production dependencies, verify claims, and coordinate exchanges that previously required more costly intermediaries. These capabilities could support reciprocal provision, distributed settlement, or revised access arrangements. Yet more detailed knowledge also makes additional transactions, finer-grained billing, and new forms of credit possible. The direction of monetary change is consequently an empirical and institutional question; richer epistemic infrastructure is not sufficient evidence of declining monetary dependence [37, 34].

The companion monetary inquiry proposed that Human–Human, Human–AI, and AI–AI relations might experience uneven or differently ordered transformations, depending in part on productive autonomy and organizational conditions. The present paper can contribute by specifying what participants would need to know and verify in order to make a particular coordination mechanism possible. It cannot determine whether those participants would consent to a clearing arrangement, whether heterogeneous values are commensurable, or whether a legally effective discharge has occurred. A physically closed exchange circuit need not be a closed monetary obligation circuit; an account balanced within one boundary may depend on excluded productive or ecological relations. Improvements in shared knowledge therefore affect some conditions of coordination without erasing its political-economic and normative dimensions.

A bounded experimental design might place a synthetic obligation ledger alongside versioned production and energy records in a federated scientific environment. The trial would vary data quality, observation latency, provenance completeness, institutional eligibility for clearing, and participant capacities to contest records. Outcomes would distinguish (i) information relevant to recognizing a possible multilateral cycle, (ii) an authorized, enacted claim reduction, (iii) monetary obligations shifted upstream, and (iv) newly created credit or access charges. A comparison group using identical production capacities with different ownership and governance rules would make institutional dependence visible. Such a trial would *not* imply that public scientific observations authorize disclosure of private commercial or personal claims.

Formally, let $\mathcal{K}_{q,h}$ denote the bounded knowledge and verification capability relevant to a coordination task q at a represented history h , and $A_{q,h}$ the set of institutionally admissible settlement actions. A candidate change in knowledge may alter which actions can be identified or evaluated, while the set of authorized actions remains separately governed. Hence

$$\Delta\mathcal{K}_{q,h} \neq 0 \not\Rightarrow \Delta A_{q,h} \neq 0 \text{ or } \Delta M_{q,h} < 0, \quad (63)$$

where $M_{q,h}$ denotes a task-specific measure of monetary dependence. Equation (63) is a logical warning about the absence of an implication, not an estimated effect. Even where epistemic improvements enable a particular clearing action, the retained claims, productive obligations, credit formation, and institutional implementation must still be evaluated. This connects the preceding research to a larger political-economic inquiry without importing the latter’s conclusions into the present paper.

Table 24 distinguishes research outputs whose differences are sometimes concealed by the general vocabulary of infrastructure. The scientific results of a marine observation mission cannot be judged solely by the operational costs of computing; a trust arrangement cannot be validated by the absence of technical errors; and a nonmonetary access service cannot be described as system-wide monetary dissolution without examining its upstream obligations. Cross-programme learning should therefore proceed through shared, contestable records and explicit translations between evaluation criteria, not through a presumed universal measure of public value.

Together, these continuations also suggest a disciplined development strategy. Begin with narrowly bounded and reversible inquiries; state the epistemic and institutional assumptions; preserve heterogeneous records and uncertainty; admit adversarial failures; document downstream dependencies and affected interests; and revise the research programme when the evidence challenges its architecture. Such

Table 24: Research continuations distinguished by their primary object, evidence, and failure tests. All listed designs are proposals; none is a completed experimental result.

Continuation	Principal research evidence	A revealing adverse result
Federated knowledge architecture	Versioned cross-node dossiers, corrected outputs, reproducibility and access tests	Records circulate correctly but downstream inferences cannot be independently revised
Ocean exploration and robotics	Calibrated marine observations, route and habitat evidence, recovery and ecological assessments	Adaptive sampling improves coverage yet systematically misses an ecologically important regime
Ecologically compatible AI infrastructure	Bounded energy, water, material, land and ecosystem accounts; comparative scenarios	Local energy savings coexist with relocated ecological harm or hidden supply dependence
Generative trust and governance	Commitment histories, dispute handling, participation, enacted remedies and exit	Formally compliant commitments provide no practical challenge or remedy
Knowledge creation and translation	Cross-context mediation dossiers, research reproduction, loss records and review histories	Fluent exchange masks incommensurability or leaves participants unable to create or contest knowledge
Epistemic infrastructure and money	Synthetic ledgers, production provenance, authorization and credit records	Improved information increases billing or credit without reducing systemic monetary dependence

an approach accommodates cooperation without assuming that national competition, institutional disagreement, or ecological constraints will disappear. It also maintains a distinction between the existence of a technical possibility and its ethical or political desirability. The broader aspiration is to expand the capacity for collective inquiry while preserving the conditions under which different participants can contribute to, evaluate, and reshape that inquiry.

14 Conclusion

This inquiry began with a practical observation and a conceptual difficulty. International scientific networks already exchange consequential measurements across jurisdictions that retain separate institutions, interests, and capacities. Atmospheric and oceanographic observations, common data policies, calibration procedures, and shared information services demonstrate that some forms of sustained scientific coordination are possible without a single authority controlling all instruments or records. The difficulty is that a network able to exchange measurements is not necessarily a network able to establish credible, revisable, and legitimately usable knowledge. The paper has consequently examined global observation and knowledge infrastructures as configurations of observational, computational, epistemic, and institutional relations. Its objective has been to specify what a defensible infrastructure inquiry must represent and test, while keeping the difference between demonstrated arrangements and proposed constructions visible throughout.

A first result is conceptual. The terms *global*, *observation*, and *knowledge* designate different properties that cannot be inferred from one another. A dataset may have international reach yet contain systematic geographic gaps; measurements may be discoverable yet inaccessible under applicable rights; systems may exchange syntactically valid records without agreeing about the quantity measured; and compatible scientific claims may be used within institutions that offer affected participants little opportunity to challenge their consequences. In the vocabulary established in Sections 1 and 2, observational coverage, access, interoperability, inferential credibility, and legitimate cooperation must be examined as related but separately evidenced achievements. Their distinction is an analytical safeguard against equating an expanding data catalogue with an expanding public capacity to know.

The meteorological and ocean-observing cases provide instructive, bounded grounds for this distinction. WMO information exchange, including WIS2 and differentiated obligations under the Unified Data Policy, demonstrates a specific way of connecting national and institutional data responsibilities through common technical and policy arrangements. Argo demonstrates a scientific circulation in which observations are published and subsequently revised through distinct quality-control processes. GOOS and associated ocean-observing coordination connect observing requirements, platforms, metadata, and operational capacities. EDITO illustrates a developing effort to integrate ocean observations, models, and applications, while remaining a programme whose implementation and performance need continuing assessment [63, 66, 55, 58, 20]. These examples provide neither a universal institutional template nor a mandate to open every category of information. Their importance lies in the mechanisms they make observable: differentiated responsibilities, standards, maintenance, quality review, and repeated cooperation under practical constraints.

A second result concerns the relational conditions of observation. A measurement becomes scientifically interpretable through a history that may include instrument configuration, sampling support, calibration, processing, inferred uncertainty, and subsequent correction. Records cannot be detached from these conditions without risking changes of meaning. The formal observation vocabulary introduced in Section 4 represents these dependencies explicitly and makes missingness, partial coverage, and version history objects of investigation. The resulting non-identifiability examples show why apparently precise outputs may fail to determine the environmental quantities that users hope to infer. An archive that preserves a numerical observation while obscuring its measurement circumstances may distribute information efficiently while diminishing a participant's capacity to assess it. Conversely, retaining provenance and uncertainty does not by itself prove scientific validity; it improves the conditions for checking particular claims.

A third result concerns the reach and limits of interoperability. Common identifiers, machine-readable metadata, shared vocabularies, and documented interfaces enable important forms of technical cooperation. The architecture proposed in Section 5 therefore includes independently governed observation nodes, federated discovery, persistent provenance, rule-governed access, versioned models, and mechanisms for correction. Yet matching metadata does not make different observation procedures identical, and a successful mapping between two representations need not preserve the inferences relevant to a particular scientific question. The formal distinctions of Section 6 treat translation as task-relative and potentially lossy; some information required for an inference may already have been discarded by an upstream transformation. This suggests that an epistemically responsible federation should record the purpose, assumptions, and limitations of its translations and preserve the possibility of principled non-equivalence. Scientific disagreement is then an object for comparison, independent checking, and reasoned revision, rather than a defect to be suppressed by a universal schema.

The assimilation inquiry reinforces this point. Multiple observations can improve inference when their errors, supports, and model relations are adequately specified; they can also create unwarranted confidence when shared biases or correlations are overlooked. Agreements among assimilated products need not imply agreement about structural models or their admissible continuations. Scientific validation therefore requires records of model choice, uncertainty characterization, independent evaluation, and changing observational conditions. The illustrative Bayesian and filtering calculations of Section 7 demonstrate logical limitations under stated assumptions rather than providing new estimates of the performance of deployed weather or ocean models. A prospective federation would need to evaluate these limitations with domain-specific data and expertise, not merely certify that participating systems produce outputs in a common format.

A fourth result is institutional. Shared interfaces do not guarantee distributed power, and a formally federated system may still be vulnerable to the loss or refusal of a critical registry, identity service, computing provider, or governance intermediary. The service-network proposition in Section 12 makes a limited version of this dependence explicit. At the same time, the alternative of concentrating every authority in one global operator introduces its own accountability and participation questions. Institutional plurality is consequently a design property to be examined through operational alternatives, rights, public responsibilities, and the practical ability to leave or contest an arrangement. The architecture proposed here does not prescribe that all participating communities accept an identical form of ownership, scientific authority, or legal organization.

This institutional concern also changes the meaning of trust. A technically authenticated message can be erroneous; a traceable decision can be harmful; and an institution may fulfil a reporting requirement while failing to correct a disputed output. Trustworthy cooperation therefore involves more than agreement on protocols or confidence in a system's numerical accuracy. It can depend on historically demonstrated performance, acknowledgment of mistakes, reciprocal commitments, meaningful participation, contestable decisions, and available remedies. In the discussion of Section 8, these features are represented as matters requiring evidence and practical institutional capacity. They should not be collapsed into a universal trust score or assumed to arise automatically from cryptographic verification. The proposed companion inquiry into generative trust can investigate how such relations form, transform, deteriorate, and become repairable without allowing this paper's architecture to determine its conclusions in advance.

A fifth result is ethical and ecological. The public purposes of scientific inquiry coexist with differentiated data rights, community authority, personal or ecological sensitivity, and unequal capacities to interpret and benefit from research. The FAIR and CARE frameworks illuminate different, complementary aspects of these concerns; neither licenses a claim that all data should be released without restriction [61, 6]. Moreover, restrictions on raw records may be ineffective where trained models or derivative products disclose sensitive attributes. Scientific institutions must therefore consider downstream inference, access to computational capacity, and the possibility of correction and remedy, including for people and communities who do not operate the underlying technical systems. These conditions are especially important as AI becomes more capable of integrating heterogeneous observations and producing persua-

sive representations whose authority may exceed their evidential support.

A further result concerns the constructive requirements of shared knowledge production and circulation. Beyond federated storage and scientific retrieval, a public infrastructure may need collaborative workspaces, research instruments, accessible computation, methodological education, recordable research histories, and mechanisms for independent correction. Moreover, generalized translation across heterogeneous GR contexts may require the situated reconstruction, negotiation, or co-creation of meanings, practices, and values. Partial mediation records can make the intended preservation criteria, translation losses, relevant authority and resulting dependencies inspectable; they cannot compel equivalent interpretations or supply legal validity and ethical legitimacy. Section 10 develops prospective services and exposes their possible adverse outcomes. Its companion research direction would construct and compare such services rather than assume a universal translation mechanism.

The infrastructure's environmental relations also remain within the research boundary. Observation platforms, data transfer, storage, model computation, maintenance, energy provision, and hardware production all rely on material capacities with potentially uneven ecological impacts. A deployment that lowers one site's operating expenses may relocate energy demand, material extraction, water use, or habitat disturbance elsewhere. No accounting method introduced in this paper is sufficient to establish that a particular proposed installation is ecologically beneficial. A responsible research design must begin with the relevant physical and ecological boundary, baseline evidence, plausible alternatives, uncertain effects, applicable responsibilities, and monitoring and remedy capacities. These requirements link scientific knowledge infrastructure to environmental research while leaving site-specific life-cycle assessment, marine ecology, and engineering validation to independent investigations.

A further connection concerns epistemic bias and the practical need for common norms. Different relationally situated systems may produce divergent understandings of the same evidence and disagreement over its permissible uses. The possibility of a bounded common protocol does not require interpretive uniformity: institutions may coordinate a specified action while preserving distinct reasons and reservations. This is a long-studied possibility in diplomacy, political philosophy, and jurisprudence, not a novelty claimed by the paper. Our narrower proposal is to make the production, interpretation, recognition, contestation and amendment of such protocols examinable as part of infrastructure governance. Agreement among present participants must not substitute for legal authority, inclusion of affected parties, or independent ethical assessment. The full theory of normative and trust formation is deliberately left to a companion study.

The hypothetical scenarios of Section 11 provide a method for challenging, rather than celebrating, the architecture. A drifting sensor can defeat apparently stable data integration; a shared vocabulary can obscure a consequential model disagreement; formally open records can be unusable to institutions lacking computational capacity; and a valid access restriction can be undermined by later publication of derived information. Each example isolates a failure that the existence of common technical interfaces cannot resolve. The examples are stipulated constructions, with no claim that their numerical settings reproduce an operational global federation. Their methodological value lies in making the proposed programme answerable to adverse observations: one should be able to identify which claim fails, which evidence is missing, and which technical or institutional revision would be required.

The research continuations in Section 13 follow from these limits without forming a compulsory technological sequence. Global knowledge architecture needs reproducible federation trials and explicit tests of provenance, authorization, and correction. Ocean exploration and autonomous robotics deserve independent scientific attention because significant marine phenomena and ecological relations remain insufficiently observed; data centres are at most one possible application of marine engineering. Ecologically compatible AI infrastructure requires material and institutional accounting, not simply claims of software autonomy. Generative trust and transnational governance require observation of commitment histories, disagreement, asymmetry, and remedies. The earlier study of monetary mediation provides a further connection: improved observation and inference may facilitate distributed coordination or clearing, but may equally support new forms of pricing, financing, and credit [37, 35]. Each inquiry has its own empirical objects and criteria of validity. Their productive relation lies in exchangeable, criticizable

methods and evidence rather than an assertion that one theory already establishes the others.

The comparison between meteorological cooperation and more heterogeneous fields also highlights the historically achieved nature of common scientific language. The stability of a formal identity under agreed axioms does not make its notation or social practice ontologically primitive. Likewise, an observed physical regularity can support powerful cross-institutional tests without proving that every possible observer must represent it through the same mathematics. By treating ontological, relational, and manifestational reality as coupled, non-hierarchical registers, the present analysis distinguishes potentially observer-independent mathematical or physical structures from the relational emergence of concepts and scientific practices and from their specific manifestations. It suspends a final account of the first register, rather than denying its possible independent reality; the principal comparison among human and hypothetical nonhuman GR systems concerns their second and third registers. A productive global knowledge infrastructure would support task-bounded common languages, document their formation and revisions, and recognize conditions under which shared manifestations conceal different generative structures.

The contribution of the present paper is consequently a bounded conceptual and architectural specification. It identifies distinct levels of interoperability, treats the measurement and revision histories of knowledge as infrastructure requirements, describes a possible federation of technical and institutional capabilities, and formulates failures that a later implementation would have to withstand. These contributions should not be mistaken for a deployed international system, a validated universal knowledge representation, an enacted transnational governance settlement, or evidence that all participating institutions would benefit equally. They remain preliminary until independently tested through comparative observation networks, reproducible computational experiments, and institutional and ecological research attentive to those affected by their results.

The central possibility investigated here is that global scientific cooperation can grow through distributed capacities for observing, interpreting, correcting, and governing shared knowledge, even where participants retain different interests, resources, and forms of authority. Its realization would depend on continued technical work and on relations of recognition, responsibility, and trust that cannot be supplied by the architecture alone. The task is therefore to construct knowledge infrastructures whose claims remain open to examination and revision, whose participation and exclusions can be challenged, and whose material conditions and ecological effects are not hidden behind the circulation of data. Such an inquiry offers a practical orientation for collective scientific work without treating the world's plurality, uncertainty, or continuing transformation as problems that a single infrastructure could finally eliminate.

A Formal Vocabulary for Observation Nodes and Histories

This appendix provides a typed, deliberately limited mathematical vocabulary for the relational and federated infrastructures examined in the main text. The purpose is to make explicit the objects invoked when the paper speaks of observation events, scientific records, enduring participating organizations, dependencies, revisions, and admissible continuations. In particular, the vocabulary separates a recorded event from a durable observer, a traceable derivation from a valid inference, and a readable record from an authorized use. It supplies formal working definitions and small logical results rather than a claim that an implemented world-scale system has been demonstrated. Sections 4 and 5 motivate its observation and architectural components; Sections 6, 7, and 8 address distinct issues of interpretation, assimilation, and institutional authority. In the author's related exploratory work, relational histories and generative representations are proposed as ways of describing knowledge formation; the present definitions are a task-specific application of that perspective [35, 34].

A.1 Types, Events, and Represented Histories

Let Evt be a type of *represented event occurrences*; its elements are not presumed to be enduring people, instruments, places, institutions, or AI systems. A chosen finite research dossier consists of a set $E \subset \text{Evt}$

and a family of typed relations among its elements. We distinguish event kinds by a type map

$$\tau : E \longrightarrow \mathcal{T}, \quad \mathcal{T} = \{\text{measure, calibrate, record, review, infer, authorize, revise, act}\}. \quad (64)$$

Equation (64) is an illustrative finite type list, extensible when a field programme distinguishes further operations. For example, a buoy’s acquisition of a salinity signal, a laboratory’s decision to revise a calibration factor, and a repository’s authorization of a derivative product are three events of different types. The buoy, laboratory, and repository are identifiable through their histories and role relations rather than being identical to those event vertices. Events can carry identifiers and observations, but an event identifier is not a metaphysical assertion of a universally unique occurrence independent of the conventions used to individuate it.

A represented history is specified as

$$\mathcal{H} = (E, D, P, L, \xi), \quad D, P \subseteq E \times E, \quad (65)$$

where D records declared epistemic or operational *dependency*, P records an asserted partial ordering under stated evidence and conventions, L is a family of additional typed relations (including observational, provenance, and governance relations), and ξ supplies evidence and uncertainty attached to those assertions. In Equation (65), D and P are separately typed: a model may depend on an earlier published result without establishing a general ordering of the physical processes described by the two sources. When P is taken to be a partial order, reflexivity, antisymmetry, and transitivity are stipulated over the represented domain, not derived from the use of a timestamp. A field-specific dependency graph need not be acyclic without qualification: repeated iterative updates can form feedback cycles. A provenance subgraph of *immutable recorded version creation* may instead be required to be acyclic under an explicit version-ranking convention.

The construction describes a *representation of an inquiry*, not a universal ontology of historical time. An investigator may compare UTC, mission time, depth-referenced geological strata, and local sampling sequences only where defensible alignment information exists. The distinction between event type, attributed order, and modelled history is essential when several institutions disagree about measurements or when the events they study have no common coordinate system.

A.2 Observation Occurrences and Record Versions

An observation occurrence $e \in E_{\text{measure}}$ is associated with zero or more *record versions*. The distinction matters because an instrument can operate while transmission fails, and because several quality-controlled products can derive from one acquisition. Let O be a set of immutable record-version identifiers and define a recorded relation $\text{Rec} \subseteq E_{\text{measure}} \times O$. For each $o \in O$, attach a typed dossier

$$\rho(o) = (\iota, y, \mathcal{S}, \kappa, \pi, v, U, P_o, \alpha). \quad (66)$$

The fields in Equation (66) correspond to the observation-record schema in Equation (6): identifier; represented signal; measurement support; calibration context; protocol; units and semantic conventions; uncertainty and quality; provenance pointers; and conditions of access and authority. Here P_o denotes a record’s provenance links, not the ordering relation P in Equation (65). The dossier is typed but need not be public in full: a sensitive species location or a restricted cultural observation may be represented by a governed pointer, access predicate, or suitably protected surrogate instead of an openly released value.

The observational support $\mathcal{S}$ is not automatically a point in a common Euclidean coordinate system. For a remote-sensing swath it may be a footprint, for an Argo profile a depth-resolved sampling path with instrument-specific reference assumptions, and for an ecological report a region deliberately generalized to protect sensitive locations. To compare records o_i, o_j , an analysis must state a scientifically relevant compatibility relation $o_i \sim_{q,c} o_j$, where q is a declared inference task and c a context of admissible alignments. This relation need not be transitive across heterogeneous methods, and the symbol $\sim$ should not be read as an equivalence relation unless the required properties have been established.

A compact schema should also allow *absent data* to be represented explicitly. Define a status map $\sigma : O \cup M \rightarrow \mathcal{S}_{\text{status}}$, where M is a set of expected measurement opportunities, and where permissible status codes differentiate unattempted collection, failed acquisition, transmission loss, unusable quality, restricted record, and model-imputed value. These labels can be coarsened for publication where revealing a detailed missingness reason would itself be sensitive. In particular, a model-generated imputation cannot acquire the type measure simply because it uses the same numerical field as an instrument product.

A.3 Nodes, Durable Roles, and Institutional Identity

A federation node is treated as a *represented role-bearing service arrangement*, not as a primitive event. Let N be a set of such arrangements with an attribution relation $\text{Operates} \subseteq N \times E \times \mathcal{R}$; its third component denotes a role such as measurement operator, data steward, calibration authority, repository, model maintainer, or dispute reviewer. For $n \in N$, the associated support of attributed events is

$$\text{supp}_E(n; r) = \{e \in E : (n, e, r) \in \text{Operates}\}, \quad r \in \mathcal{R}. \quad (67)$$

The map in Equation (67) permits one institution to play several roles and multiple institutions to participate in a single event. An organizational succession may replace a node identifier or redistribute roles while scientific records remain citable. Neither the existence of a software endpoint nor a signed data product establishes legal personhood, ownership, or moral standing. Those questions require distinct institutional and jurisdictional descriptions, especially where an AI-operated service acts through a responsible human or corporate organization.

Durable identity across a changing service configuration is necessarily evidence-relative. A persistent identifier may link versions of a buoy, operator, or repository despite hardware replacement; such persistence is not evidence that every function, calibration chain, or authorization survives unchanged. One can model identity hypotheses as links between subhistories supported by maintenance logs, continuity of responsibility, or governance records. Rival continuity claims may coexist pending adjudication. This is why the architecture's ability to resolve a resource identifier is separate from the question of who can legitimately bind the associated institution to a commitment.

A.4 Coordinate Charts and Partial Synchronization

For an observation programme a , let $\chi_a : E_a \rightarrow C_a$ be a partial map assigning locally justified coordinates to selected events. A chart C_a may combine spatial support, local time labels, mission sequence, depth, or other task-dependent indices. For programmes a, b , coordinate translation is a *partial, evidenced* map $\phi_{ab} : D_{ab} \subseteq C_a \rightarrow C_b$; it carries an uncertainty or qualification record ω_{ab} . A suitable translation claim may be stated as

$$\chi_b(e) \approx_{\omega_{ab}} \phi_{ab}(\chi_a(e)), \quad e \in E_a \cap E_b, \quad \chi_a(e) \in D_{ab}, \quad (68)$$

where $\approx_{\omega_{ab}}$ means agreement under an explicitly specified tolerance, transformation, or probabilistic error model. Equation (68) neither asserts that the same event appears in every chart nor guarantees that mappings between three charts compose without loss. Approximate pairwise consistency can fail around a cycle of translations, and incompatible sampling supports can defeat a numerical timestamp match. Any putative global spatiotemporal coordinate representation therefore remains a derived, revisable modelling achievement.

In practical ocean–atmosphere comparisons, this discipline protects distinctions among an in situ measurement interval, a satellite pixel aggregation, and the support of an assimilated model cell. Some correlations can be estimated without asserting that the records refer to identical sampling supports. Likewise, a causal or informational dependency edge should not be inferred simply from an earlier timestamp. The edge requires evidence of source use, operational procedure, or an argued causal model.

A.5 Pathwise Consistency of Partial Coordinate Translations

A collection of individually well-formed translation functions need not yield the same result along alternative paths. This is an issue of documented compatibility among representations, rather than a claim that there exists a unique global spacetime chart. For maps ϕ_{ab} , ϕ_{bc} and ϕ_{ac} , equality of the direct and composed path requires the common domain $D_{abc} = \{x \in D_{ab} : \phi_{ab}(x) \in D_{bc}\} \cap D_{ac}$ and an additional compatibility assertion, for example

$$d_c(\phi_{bc}(\phi_{ab}(x)), \phi_{ac}(x)) \leq \eta(x), \quad x \in D_{abc}, \quad (69)$$

where d_c is a justified comparison metric in the chart C_c and η is a declared path-discrepancy tolerance. The expression concerns the agreement of two representations under an agreed evaluation task; it neither creates a global coordinate system nor makes the indicated path metaphysically preferred.

Proposition A.1 (Individual map definition is insufficient for path agreement). *Even with total functions between identical coordinate sets, the existence of ϕ_{ab} , ϕ_{bc} and ϕ_{ac} does not entail $\phi_{bc} \circ \phi_{ab} = \phi_{ac}$.*

Proof. Let each coordinate set be $\mathbb{R}$, with $\phi_{ab}(x) = x$, $\phi_{bc}(x) = x$, and $\phi_{ac}(x) = x + 1$. Each map is total and well defined, yet $\phi_{bc}(\phi_{ab}(x)) = x \neq x + 1 = \phi_{ac}(x)$ for every x . $\square$

The counterexample isolates a missing *coherence condition*: it does not assert that these three maps would pass empirical calibration against the same events. In a deployed observatory, map versions, common sampling support, path-specific uncertainty, and the institution responsible for each mapping need to accompany a claim of compositional adequacy. If two paths fail to agree within a justified tolerance, retaining their competing results and initiating review is more defensible than silently selecting the shortest available path. The same qualification applies to generalized interpretive translations, although their agreement criteria need not be numeric.

A.6 Typed Dependencies and Provenance Subgraphs

Let $\mathcal{K} = \{\text{uses, calibrates, derives, reviews, supersedes, authorizes, challenges}\}$ denote an illustrative set of edge types. Then a typed dependency multigraph is $G_{\mathcal{H}} = (V, A, s, t, k)$, where $V = E \sqcup O \sqcup N$ is a disjoint typed union, A is a set of relation assertions, $s, t : A \rightarrow V$ identify their sources and targets, and $k : A \rightarrow \mathcal{K}$ assigns edge kinds. This structure can represent a correction, a model's input dependency, and an authorization as different kinds of edges rather than treating all links as a single undifferentiated provenance relation. Such typing can be exported to a more elaborate provenance vocabulary, but the present mathematical definition is not a formal implementation of W3C PROV [71].

An immutable *version-creation* subgraph G_V is obtained by retaining derives and supersedes edges that point from an existing record version toward a newly created version. Let $h : O \rightarrow \mathbb{N}$ be a verifiable creation-rank function on the relevant dossier. We impose

$$(o_i, o_j) \in A_V \implies h(o_i) < h(o_j). \quad (70)$$

Proposition A.2 (Acyclicity of rank-respecting recorded versions). *For a finite version graph G_V satisfying Equation (70), G_V contains no directed cycle.*

Proof. A directed cycle would yield $h(o_1) < h(o_2) < \dots < h(o_m) < h(o_1)$ by repeated application of Equation (70). This contradicts the strict ordering of integers. $\square$

Proposition A.2 is conditional on a trusted creation ranking; it does not establish that signatures are authentic, that records are truthful, that the claimed derivation occurred, or that scientific dependency itself is acyclic. For instance, numerical models may incorporate iterative feedback, and rival laboratories may challenge one another in a cycle of response. A system can preserve both: an acyclic log of immutable assertions concerning a potentially cyclic scientific debate. Nor does a corrected product erase the previous version; the historical chain preserves the original assertion and the explicit reasons for later revision. This distinction is consistent with the paper's emphasis on reconstructable quality-control and correction histories [56, 71].

A.7 Scientific Claims, Identifiability, and Task Relative Inference

It is useful to separate measurement products from the scientific claims derived from them. Let X be a chosen space of candidate model states or histories, Y a space of records admissible under a specified observation design, and $F : X \rightarrow Y$ a deterministic observation map. Let $q : X \rightarrow Z$ denote a target scientific quantity or claim. An exact record-based inference rule for q is a function $g : Y \rightarrow Z$ satisfying $q = g \circ F$. Its existence is a *task-relative* form of identifiability, not a proof that the complete model state is observable.

Proposition A.3 (Task-relative factorization criterion). *Let $F : X \rightarrow Y$ and $q : X \rightarrow Z$ be functions. There exists a function $g : F(X) \rightarrow Z$ such that $q = g \circ F$ if and only if $q(x_1) = q(x_2)$ whenever $F(x_1) = F(x_2)$.*

Proof. If $q = g \circ F$, equality of the observations forces equality of the target values. Conversely, assume q is constant on every fiber $F^{-1}(y)$ for $y \in F(X)$. Define $g(y) = q(x)$ using any x satisfying $F(x) = y$. Constancy on the fiber makes g well defined and gives $q = g \circ F$. $\square$

In a deliberately simplified coastal example, suppose $x = (x_1, x_2)$ represents contributions from two locations, while a coarse sensor records only $F(x) = x_1 + x_2$. The total $q_+(x) = x_1 + x_2$ factors through F exactly, whereas the contrast $q_-(x) = x_1 - x_2$ does not. States $(1, 0)$ and $(0, 1)$ yield the same observation but different contrasts. Proposition A.3 formalizes the point that a dataset may be fully adequate for one inquiry and irretrievably insufficient for another. The result is distinct from an assessment of noisy inference, probabilistic uncertainty, or the social authority of an interpreted claim; those require additional modelling assumptions. It also supports the main text's caution against treating semantic agreement and model translation as a general guarantee of inferential equivalence (Section 6).

A.8 Permissions, Responsibility, and Noninheritance of Access

Institutional authority introduces a layer that cannot be recovered from scientific provenance alone. For an actor or attributed role a , a record version o , an intended purpose p , and a context c (including jurisdiction, affected rights, emergency conditions, and current policy versions), define a partial decision function

$$\text{Auth}(a, o, p, c) \in \{\text{permit}, \text{deny}, \text{undetermined}\}. \quad (71)$$

The third result in Equation (71) indicates a decision not established by the available governance records; it is not permission. A successful metadata lookup or valid cryptographic signature does not alter this outcome by itself. The authorization relation may also depend on transformed outputs, identifiable communities, protected ecological sites, downstream use, and dispute status; those requirements should be made explicit in a deployed scheme [6, 57].

Proposition A.4 (Permissions do not follow derivation by logic alone). *The existence of a derivation edge $o \rightarrow o'$ and permission to access o does not entail permission to access o' unless an additional governance rule establishes that implication.*

Proof. Consider a governance context in which access to source record o is permitted for quality assessment but a derived product o' reveals a protected habitat location and is restricted. Set $\text{Auth}(a, o, p, c) = \text{permit}$ and $\text{Auth}(a, o', p, c) = \text{deny}$. These decisions are logically consistent with the derivation edge. Hence the implication does not follow from provenance alone. $\square$

Proposition A.4 is a statement about underdetermination by the chosen representation, not a jurisdiction-specific legal ruling. The converse implication also fails generally: access to a public aggregate does not grant access to its protected components. Restrictions applied to original files may be insufficient if model outputs allow sensitive content to be reconstructed. Conversely, retaining a restricted input identifier may enable authorized audit without publishing the underlying record. The design requirement is therefore to represent permissions and their justifications alongside, but independently of, provenance and inference relations.

A.9 Continuations, Revisions, and Representational Commitments

Let $\mathfrak{H}$ be a class of well-typed represented histories. Given $\mathcal{H} \in \mathfrak{H}$, the set $\mathcal{C}(\mathcal{H})$ of *admissible continuations* contains diagrams of additional, corrected, or reorganized records and relations that satisfy explicitly declared consistency and governance constraints. We use the word continuation for a permissible extension or transformation of a representation, not for a metaphysically privileged future of the underlying world. When new evidence adds a previously unrecorded sample, an inclusion $\mathcal{H} \hookrightarrow \mathcal{H}'$ may suffice. When a scientific interpretation is revised, a span

$$\mathcal{H} \xleftarrow{j} \mathcal{K} \xrightarrow{j'} \mathcal{H}' \quad (72)$$

can identify a common auditable part $\mathcal{K}$ without requiring the old and new scientific claims to be monotone refinements of the same truth assignment. Equation (72) is a diagrammatic convention: it neither requires a category of all histories to be constructed in this paper nor licenses treating corrections as deletion of historical evidence.

For practical provenance, an append-only event log can record the creation of revised statements and new evidence while a user-facing *current view* changes nonmonotonically. Suppose a profile's salinity value is adjusted after delayed-mode quality control. The previously released product remains recoverable under its version identifier, the later product records its calibration basis, and a model that used the older version can remain reproducible. The most recent record is not automatically the most accurate for every purpose; quality assessment requires declared measurands, validation procedures, and uncertainty descriptions.

A continuation relation can also model withdrawal of access or the dissolution of an organizational partnership. Deleting access from a current authorization view does not retroactively guarantee that already disseminated copies have become inaccessible. A rigorous federation must distinguish a revised policy, attempted enforcement, reported dissemination history, and actual use. Consequently, revision, authorization, and scientific quality are separate predicates over a history, even where their records are stored within one service.

A.10 Worked Hypothetical Federation Dossier

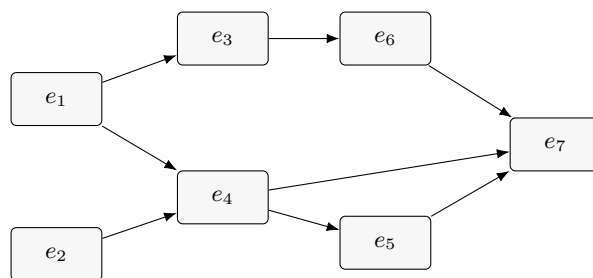
Consider a purely hypothetical coastal study with three role-bearing nodes: n_B (a buoy operator), n_S (a satellite data steward), and n_L (a scientific laboratory). We stipulate seven represented events: e_1 buoy acquisition; e_2 satellite acquisition; e_3 laboratory calibration review; e_4 preliminary assimilation; e_5 external validation; e_6 revised buoy product; and e_7 revised scientific assessment. The scientific dependencies are declared as $e_1 \rightarrow e_3$, $e_1 \rightarrow e_4$, $e_2 \rightarrow e_4$, $e_3 \rightarrow e_6$, $e_4 \rightarrow e_5$, $e_4 \rightarrow e_7$, $e_5 \rightarrow e_7$, and $e_6 \rightarrow e_7$. They are asserted *uses or reviews* relations, not an ontological universal order. In particular, e_2 and e_3 need not be ordered relative to one another; the dossier records only their relationships to downstream operations.

Table 25: Hypothetical events, records, and roles in a federated coastal inquiry.

Event	Declared operation	Role-bearing node	Recorded epistemic status
e_1	In situ acquisition	n_B	Buoy signal with uncertain drift
e_2	Remote acquisition	n_S	Satellite product with distinct support
e_3	Calibration review	n_L	Candidate offset and rationale
e_4	Preliminary assimilation	n_L	Model-based synthesis using initial versions
e_5	External validation	n_S, n_L	Comparison and documented disagreement
e_6	Revised buoy product	n_B, n_L	Adjusted value with new version identifier
e_7	Revised scientific assessment	n_L	Recomputed claim with provenance and uncertainty

Table 25 contains no empirical measurements. It is a type-checking example. A graph of the declared

dependencies is shown in Figure 3; no edge is intended to mean that an institution, sensor, or AI system is a primitive vertex of spacetime.



Arrows: stipulated epistemic or operational dependencies

Figure 3: A finite hypothetical dependency diagram. Vertices are represented events; the nodes that operate the federation are role attributions, not the vertices.

Suppose the laboratory discovers a drift correction after the initial assimilation. It publishes the revised buoy record as a new immutable version and cites the original as its source; it does not silently replace e_1 . The revised assessment e_7 may then depend on both the original model product and the revised calibration record. If the satellite-derived observations have a distinct spatial footprint, mathematical co-registration must be stated rather than presumed. If the source includes a sensitive ecological record, authorization for that record and any derived product must be evaluated separately. Different reviewers may reach divergent assessments on the same data, which the dossier can record without equating their disagreement with a technical failure.

The example exposes four different tests: whether the dependency assertions are traceable; whether the measurements and uncertainties support the scientific claim; whether participating nodes have adequate operational roles and capacities; and whether proposed access or action is legitimately authorized. None follows from the other three by definition.

A.11 Scope, Failure Conditions, and Further Formal Development

The preceding construction gives a type discipline useful for this discussion paper, but leaves several objects intentionally open. Events are selected and individuated by research conventions; a history's recorded ordering may be incomplete or contested; scientific dependency may contain feedback; and reliable version logs depend on trustworthy operational practices. Measurement values and calibration contexts require domain-specific scientific validation. Coordinate alignments are partial and uncertainty-bearing. Institutional authority depends on real rules and affected parties. None of these features can be guaranteed by a general graph vocabulary.

Several further results would require additional assumptions or evidence. An observability theorem for a deployed ocean or atmospheric system needs a validated physical state model and calibrated observation operations. Statistical claims about missingness or forecast performance need sampling and held-out data. A proof of robust interinstitutional correction needs a protocol model, failure assumptions, and specified decision rights. Evaluating actual data-access safeguards requires security and rights analysis. Ecological and energy impacts require material measurements and explicit boundaries. These tasks are intentionally separated from the elementary implications proved in Propositions A.2, A.3, and A.4.

The formal vocabulary will support the later appendices' treatment of federated metadata, translation counterexamples, illustrative cross-border networks, governance disputes, and evaluation protocols. Its central methodological claim is modest: observations, interpretations, institutions, and historical continuations can be represented with carefully differentiated types and auditable relations, while the adequacy of those representations remains open to scientific scrutiny, institutional contestation, and revision.

B Federated Metadata and Provenance Schema

This appendix translates the typed vocabulary of Appendix A into a preliminary application profile for a federation of observation and research institutions. Its objective is to specify what a participating node could publish, validate, preserve, and contest. The proposal is deliberately *not* a claim that one universal schema can exhaust heterogeneous scientific practices. A metadata profile permits exchange of a bounded set of assertions; it does not prove the truth of those assertions, the equivalence of their interpretations, or the legitimacy of the rules through which they are disclosed. The profile separates the observation occurrence, a released record version, the scientific operation applied to a record, its attributed institutional responsibility, and the decision governing a particular use. These are related objects, rather than competing names for a single object.

The design draws selectively on published vocabularies and technical specifications: W3C PROV for provenance relations, DCAT 3 for catalogues and dataset distributions, DataCite 4.7 for identifiers and citation metadata, OGC API-Records for discovery interfaces, SOSA/SSN for observation semantics, and CF conventions for scientific variables and coordinate descriptions. JSON Schema 2020-12 supplies a possible machine-readable validation language; SHACL can serve graph-oriented profiles. None of these specifications imposes the full cross-institutional governance model examined here [71, 74, 12, 50, 75, 7, 44, 72]. The names and identifiers used in the example are purely illustrative, not active data services or registered identifiers.

B.1 A Bounded Interoperability Profile

Let N denote a collection of federation participants, O immutable record versions, E represented events, and A attributed agents or institutional roles. The proposed federation profile is a tuple

$$\mathcal{F} = (\Sigma, \text{Id}, \text{Prov}, \text{Sem}, \text{Qual}, \text{Pol}, \text{Disc}, \text{Val}), \quad (73)$$

where Σ gives field types and required-field rules; Id supplies identification and version conventions; Prov records derivation and revision assertions; Sem identifies declared measurement semantics; Qual represents uncertainty and data quality; Pol links governing policies and responsible authorities; Disc defines public or restricted discovery projections; and Val defines validation procedures. Equation (73) describes a candidate application profile, not a common ontology to which every observation programme must conform internally. Implementers would map this bounded profile to existing service metadata, retaining native records and recording any lossy mappings.

The minimum interoperable unit is a *versioned assertion envelope*, rather than a complete scientific database. An envelope identifies what was published, by which attributed node and record-production process, what it purports to describe, how it can be located or requested, and the rules under which the claim may be used. It need not expose restricted measurements. A record may be discoverable by a protected surrogate that does not disclose an exact location, and an inaccessible source can remain citable through a suitable governed identifier. The federation should distinguish *discoverable*, *requestable*, *retrievable*, *reusable*, and *reproducible* as independent predicates. These distinctions extend the access, authority, and model-validation considerations developed in Sections 5 and 9.

B.2 Record Identity, Version Identity, and Fixity

A stable *resource identifier* i groups representations considered versions of a resource by an explicit stewarding convention; a *version identifier* v identifies an immutable released snapshot. Let $\text{res} : O \rightarrow I$ and $\text{ver} : O \rightarrow V$ be the corresponding functions. The pair (i, v) may address a snapshot, but resolution of i alone may select a mutable default view. A published metadata record should distinguish those operations. A persistent DOI can be useful for citation without substituting for bitstream fixity, which may be checked using a named cryptographic digest. DataCite explicitly distinguishes updates to metadata from significant content revisions in its versioning guidance [12, 13]. In this appendix, no DOI is fabricated or implied for the worked example.

For an archived byte sequence b and named digest algorithm H , an integrity assertion takes the form

$$\text{Fixity}(v, b) \iff H(b) = d_v, \quad (74)$$

where d_v is the trusted digest committed to record version v . A match in Equation (74) supports identity of a byte sequence relative to an already trusted digest and algorithm. It does not establish the authenticity of the originally supplied digest, scientific correctness, consent, or completeness of the original observations. Signatures, key management, and independent validation introduce distinct evidence. In particular, a checksum is not an environmental measurement-quality certificate.

The schema treats recorded acquisition times and publication times as attributed coordinates with declared clock, uncertainty, and calibration, rather than proofs of an underlying universal temporal order. *Supersedes*, *is derived from*, and *is corrected by* are distinct relation types. A new derived product does not necessarily supersede its inputs; a revised quality assessment may supersede an earlier assessment while citing the same original signal. Persistent identity, identity of bytes, scientific equivalence, and institutional continuity are therefore separate questions.

B.3 Typed Core Metadata and Scientific Support

Table 26 presents a compact core profile. Optionality is evaluated relative to a named resource type, so missing values can have explicit reasons; an empty string is not a satisfactory substitute for unknown calibration or intentionally withheld location. The proposed machine-readable companion implements a constrained subset of these fields. Its validation is demonstrative and must not be confused with official conformance to any of the referenced standards.

Table 26: Proposed fields in a versioned federated record envelope. The fields are locally proposed extensions or mappings, not a new international standard.

Field group	Content and intended interpretation	Risk or verification boundary
Identity and type	Stable resource ID, immutable version ID, version label, record kind, schema-profile version	Resolution and fixity do not prove the recorded event occurred
Attribution	Steward ID, operational role, asserted activity or observation-event ID	Attributed roles require independent organizational evidence
Observation semantics	Measurand URI, method URI, units, coordinate-frame references, sampling support	Shared terms do not prove calibration or scientific equivalence
Quality and status	QC procedure/version, calibration reference, uncertainty description, missingness code	Quality labels are asserted and task-relative
Provenance	Typed references to input versions, creation/review events, processing software and parameters	A lineage graph may omit relevant unrecorded dependencies
Fixity and versioning	Digest algorithm/value, supersession or revision links, publication assertion	Matching a digest establishes neither truth nor authorization
Governance	Steward, policy/version reference, permitted-purpose query, decision, appeal channel	A policy URI alone does not establish effective rights or remedy
Discovery and access	Public surrogate fields, service endpoint, access mode, protected-detail indicator	Metadata alone can leak protected facts

Observation semantics must explicitly identify the *measurand* and relevant support: an instantaneous reading at a point, a depth interval, a satellite footprint, an inferred grid cell, and a scenario-model output are not interchangeable. When two sources report the same variable name, comparison also depends on sampling and aggregation procedures, units, frames, calibration, and uncertainty. A field may point to a vocabulary concept rather than reproduce its definition. Locally tailored mappings then retain their

derivation, loss conditions, and human review status. Applying CF or SOSA/SSN descriptions can aid such work without asserting a unique generative model of the represented environment [75, 7].

A particularly important status distinction concerns *measured*, *simulated*, *assimilated*, *imputed*, and *review-only* products. Data lineage should prevent an assimilated field from silently acquiring the status of an independent observation. Likewise, a missing record may mean that measurement was not attempted, acquisition failed, the value failed quality control, dissemination was prevented, or the data were restricted. These explanations can themselves be sensitive. Hence missingness status may have a protected detail layer with a non-disclosing public status.

B.4 Provenance Assertions, Revisions, and Knowledge Claims

A provenance assertion is represented by a typed edge (u, k, v, a, e) relating source and target record versions u, v , edge kind k , asserting role a , and recorded assertion event e . The kinds *used*, *wasGeneratedBy*, *wasDerivedFrom*, *wasAttributedTo*, and *wasRevisionOf* can be mapped as appropriate to W3C PROV. This mapping is partial and should not be advertised as lossless, since local scientific relations such as *invalidates a calibration assumption* have domain-specific meaning [71]. Let $G_P = (O, L_P)$ be the asserted record-lineage multigraph. A machine can verify well-typed endpoints and declared edge constraints without proving that every dependency actually occurred.

A record may carry a structured *claim* or interpretation, including its provenance and supporting method. We distinguish three predicates,

$$\text{SchemaValid}(o), \quad \text{Traceable}(o), \quad \text{ScientificallySupported}(o; q, c), \quad (75)$$

where the first checks the structure of the envelope, the second demands an adequately resolvable declared lineage for a chosen provenance profile, and the third expresses a task- and context-dependent epistemic judgment. The predicates in Equation (75) are independent in general. A well-formed lineage can document a scientifically mistaken calibration, while scientifically persuasive evidence can temporarily lack complete machine-readable metadata.

Proposition B.1 (Syntactic and provenance validity do not imply epistemic validity). *Even if $\text{SchemaValid}(o)$ and $\text{Traceable}(o)$ both hold under a chosen schema and complete declared provenance graph, these conditions alone do not entail $\text{ScientificallySupported}(o; q, c)$ for every task q and context c .*

Proof. Construct two well-typed records with equal permitted scientific fields and valid chains of acquisition and release. In one dossier, the instrument’s calibration is accurate; in the other, it has an unobserved additive bias. Both can have complete and consistent asserted provenance, while only the first supports a chosen high-accuracy quantitative inference. The structural and tracing predicates do not discriminate between the two unobserved calibration states, so the implication fails. $\square$

Proposition B.1 is a limitation of the asserted representation; it is not an allegation that any named programme uses incorrect calibration. A federation must permit a later scientific challenge to attach a review event and issue a new released version. The current view may change nonmonotonically as evidence evolves, even while the record of what was previously asserted remains auditable. Version selection should be a documented *policy function* of release status, validation scope, withdrawal, and intended scientific use, not simply the operation of choosing the greatest timestamp or version number.

B.5 Governance Decisions, Disclosure, and Purpose Limitation

Access control is modeled as a contextual decision over an actor or role a , a target record or product o , purpose p , governance context c , and policy version λ :

$$\text{Dec}_\lambda(a, o, p, c) \in \{\text{permit}, \text{deny}, \text{undetermined}\}. \quad (76)$$

An *undetermined* result from Equation (76) is neither presumed approval nor scientific invalidity. Institutionally binding decisions require an identified competent authority, reviewability, appeal, and enforcement. A metadata profile can preserve evidence that a decision was asserted, but it cannot create legitimate authority by technical declaration. A consent record may also be scoped to an institution, project, purpose, and product version; consent for raw data processing does not invariably authorize public release of model outputs. The design is informed by public-data stewardship and community authority principles, while remaining jurisdiction neutral [6, 57, 25].

Federated policy evaluation needs a declared combination rule. The rule *permit if any node permits* is unsafe when different nodes govern distinct resources or purposes. Conversely, *deny if any node denies* can suppress independent, legitimately public material. We propose no universal combination algorithm; instead a request should identify the resource’s relevant stewards and the applicable jurisdictional and contractual rules. A federation response should be capable of expressing a contested policy interpretation and the procedure for review. Authentication establishes a presented credential under specified trust assumptions; it does not determine whether the role holder is entitled to grant access.

A public discovery record can be viewed as a projection $\pi_P(o)$ of a richer protected record o . An apparently minimal projection can still reveal sensitive facts through a title, geographic bounding box, record count, identifier, dataset linkage, or update frequency. Therefore the safety of the public index is a property of its entire disclosure process and adversarial context, not of its adherence to a short field list.

Proposition B.2 (Field exclusion does not guarantee contextual non-disclosure). *The absence of a protected field from a public discovery projection does not generally guarantee that the protected attribute cannot be inferred from that projection and public auxiliary information.*

Proof. Consider a record about the presence of a protected species in one of two habitats. Remove the explicit species-location field, but publish a catalogue collection name that uniquely denotes one habitat. An observer with the public mapping from collection names to habitats can reconstruct the location. The protected field was excluded while its value remained inferable from the released projection and auxiliary information. Hence field exclusion alone does not imply non-disclosure. $\square$

Proposition B.2 motivates threat modelling, contextual review, and protection of derived outputs, including AI-accessible summaries and embeddings. It does not assert that all disclosure can be prevented. If an access policy changes, the system should record the policy update, affected versions and endpoints, notice to downstream stewards, and any enforceable recall or takedown process. Complete revocation of already copied information cannot be inferred from updating a discovery record.

B.6 Federated Discovery and Cross-System Validation

The federation should support a small number of declared discovery and retrieval operations: retrieve a public catalogue entry by resource identifier, query by safe metadata facets, request a named immutable version, obtain a policy decision, and retrieve an authorized provenance subset. The specific HTTP paths in a deployment would follow the chosen service standard rather than a universal prescription of this paper. OGC API-Records and DCAT 3 offer existing models for catalogue discovery and exchange of descriptions; DataCite contributes stable citation and inter-resource links [50, 74, 12]. A catalogue response should not be mistaken for evidence that a referenced payload is actually available to a given requester.

A layered validation pipeline can be summarized as

$$\text{Parse} \prec \text{Shape} \prec \text{Resolve} \prec \text{SemanticReview} \prec \text{ScientificReview} \prec \text{PolicyDecision}, \quad (77)$$

where $\prec$ represents a suggested *checking dependency* rather than mandatory serial execution. Some tasks can run concurrently; a governance decision may block further scientific processing before data are disclosed. The layers mean: parse the format; validate local schema and controlled terms; resolve claimed

identifiers and version links; check meaning, frames, and admissible translations; review calibration, uncertainty, and inference; finally adjudicate use under the relevant policy. A failure in one layer should be reported with its own evidence and status rather than overwritten by an undifferentiated “invalid” flag.

To test an interoperability mapping $T_{ab} : \Sigma_a \rightarrow \Sigma_b$, one must declare a set Q of scientific queries whose preservation matters. A proposed test is

$$\forall q \in Q, \forall o \in D_{ab} : \text{Ans}_b(q, T_{ab}(o)) \approx_{\epsilon(q,o)} \text{Ans}_a(q, o), \quad (78)$$

with explicitly specified domain D_{ab} , error model ϵ , and evidential conditions. Equation (78) does not require total or invertible translation. In particular, a coarsened spatial record can preserve regional totals while failing a habitat-level inquiry. This connects the application profile to the task-relative identifiability result in Proposition A.3 and to the main text’s treatment of epistemic translation (Section 6).

The cross-layer review profile in Equation (18) adds an important restriction to the application profile: the result of the semantic-preservation test in Equation (78) is an attributed epistemic assessment, while the result of a policy check is a separate record concerning a specified action, applicable rules and a claimed competent source. Neither a successful mapping nor a syntactically well-formed permission statement proves the other. A protocol implementation should retain a review dependency link for each assessment and distinguish missing evidence from an adverse decision. The accompanying synthetic fixture examples/fictional_cross_layer_review.json checks this type separation only; it does not establish legality or field performance.

B.7 Worked Hypothetical Revision and Restricted Derivative

Consider a fictional coastal federation with a buoy operator N_B , an ocean laboratory N_L , and a community data steward N_C . The operator issues version v_1 of a temperature profile o with a stated but subsequently questioned calibration. The laboratory produces an assimilation product m_1 using v_1 and a satellite dataset. A later review event identifies a sensor-offset problem and issues corrected profile version v_2 . The laboratory produces m_2 with revised inputs; the historical record preserves a typed link from v_2 to v_1 as a correction and from m_2 to m_1 as a later research product. A separate habitat inference h_1 derived from m_2 may reveal a sensitive breeding site. A valid lineage link to m_2 neither grants open access to h_1 nor establishes its ecological reliability.

Table 27: Hypothetical processing and governance decisions across versioned coastal records.

Version	Asserted kind and relationship	Publication context	Interpretive or governance caveat
v_1	Instrument profile; acquired under calibration k_1	Public for specified research purposes	Later challenged; source version remains citable
m_1	Assimilation product uses v_1 and another dataset	Public aggregate under documented method	Does not constitute independent measurement
v_2	Corrected profile; revision of v_1	Public with correction notice	Quality change requires supporting review
m_2	Reanalysis uses v_2 ; supersedes one public model view	Public for declared purposes	Some interpretations may remain disputed
h_1	Derived habitat estimate uses m_2	Restricted pending community and ecological assessment	Public discovery record must avoid habitat leakage

A minimal machine-readable specimen accompanies the LaTeX source in examples/, with a locally proposed JSON Schema, a fictional record, and a validation test. The files are an *application-profile sketch*, not official DataCite, DCAT, PROV, or WMO conformant encodings. They use obviously fictitious example identifiers and omit real sensitive coordinates. The script checks structural constraints and version consistency for this case, but cannot evaluate oceanographic truth, ethics, competent authority, or the sustainability of infrastructure operation.

B.8 Implementation and Evaluation Boundaries

The proposed profile can be evaluated on controlled cross-node exercises. A useful test suite would include a valid but scientifically erroneous calibration record, a scientifically credible record with missing metadata, an inaccessible source with a public citation surrogate, a changed governance rule, a noninvertible units-and-support translation, and a restricted derivative whose existence alone reveals sensitive information. The evaluation should report structural validity, discoverability, resolvability, correction latency, task-specific inferential loss, legitimate access success, unauthorized disclosure risk, and the ability of participating communities to challenge an interpretation. No single index measures all these properties.

Three practical limits remain. First, scientific fields inevitably vary between meteorological, oceanographic, genomic, ecological, and social inquiry; a global core must tolerate domain-specific profiles and acknowledge mapping loss. Second, the infrastructure depends on people and organizations maintaining sensors, vocabularies, policies, audit channels, and electricity-intensive storage and computation; a published schema does not discharge those obligations. Third, institutional disagreement can remain even when all parties exchange syntactically valid records. Consequently, this appendix proposes a testable representation for *recorded and governable epistemic relations*, not an algorithm for replacing independent scientific inquiry or legitimate collective decision-making. The later appendices develop semantic counterexamples, institutional network scenarios, evaluation metrics, and the accompanying evidence ledger.

C Semantics and Translation Counterexamples

The preceding appendices developed a vocabulary of represented observation events and an application profile for versioned records. This appendix investigates what those representations do *not* guarantee. A node may publish schema-valid data, a receiving node may parse it correctly, a vocabulary mapping may carry an `exactMatch` assertion, and the resulting exchange may nevertheless fail to preserve a scientific question. The issue is not merely erroneous conversion. A representation can discard distinctions essential to a later inquiry, a translation can change measurement support, and different generative models can remain observationally indistinguishable on the records selected for exchange. Accordingly, the counterexamples below identify the target quantity, the purported translation, the information or authority omitted, and the condition under which the problem could be detected or repaired. They concern deliberately fictional data and idealized models; they are neither demonstrations of operational failures in named international networks nor evidence that any one semantic standard is defective.

This analysis extends Section 6 and Appendix A, and is motivated in part by the earlier exploratory inquiry into inference of another system's generative representation [35]. The role of the latter is methodological: exchanging outputs can be less demanding than reproducing the processes by which outputs acquire meaning. Technical vocabularies such as SOSA/SSN, SKOS, and CF conventions help express measurement and concept distinctions, while a federation must still test which inferences survive a particular translation [75, 70, 7]. We make no assumption that a universal language of all scientific inquiries exists.

C.1 Typed Translation Claims and Their Scope

Let X_A and X_B be sets of admissible records under the native representations of participating systems A and B . A translation claim should minimally specify the tuple

$$\tau_{AB} = (D, T_{AB}, q_A, q_B, \mathcal{Y}, d, \varepsilon, \gamma), \quad (79)$$

where $D \subseteq X_A$ is the declared domain, $T_{AB} : D \rightarrow X_B$ is a translation, $q_A : X_A \rightarrow \mathcal{Y}$ and $q_B : X_B \rightarrow \mathcal{Y}$ are interpretations of a particular scientific quantity in a justified common space $\mathcal{Y}$, d is a comparison

function, ε is a tolerance or error-bound rule, and γ identifies the relevant measurement, model, version, and authorization context. The claim is *task-adequate* only if

$$d(q_A(x), q_B(T_{AB}(x))) \leq \varepsilon(x) \quad \text{for every } x \in D. \quad (80)$$

Neither Equation (79) nor Equation (80) asserts that the objects represented by A and B are identical. The requirement to specify the comparison space matters: if one system records a biodiversity count and another a legal designation of protected status, there is generally no canonical numerical d between the two without an additional operational or normative interpretation. Even when d exists, tolerances may vary by location, instrument support, decision threshold, or intended use.

A translation claim should also identify *what was not tested*. A publicly reported mean may be sufficient for a regional-average question but not for extremes; a mesh-cell estimate may be adequate for coarse forecasting but not for a habitat-localization decision; a concept mapping may be adequate for search discovery but not for automated scientific or legal substitution. The metadata profile of Appendix B can record such declarations and their versions. Its structural validation remains separate from checking Equation (80) against independently constructed cases.

C.2 Aggregation and the Impossibility of Exact Reconstruction

Consider two samples $x = (x_1, x_2) \in \mathbb{R}^2$ from a fictional environmental monitoring process. System A retains the pair, while system B stores only its arithmetic mean,

$$T(x_1, x_2) = \frac{1}{2}(x_1 + x_2). \quad (81)$$

Suppose a later study asks for an extreme-value statistic $q(x) = \max\{x_1, x_2\}$. The data vectors $x = (0, 0)$ and $x' = (-2, 2)$ produce the same translated record $T(x) = T(x') = 0$, but $q(x) = 0$ and $q(x') = 2$. The question was not preserved by aggregation, even though the mean was calculated exactly and the translated record may be perfectly standards-compliant.

Proposition C.1 (Exact target reconstruction requires fiberwise invariance). *For a deterministic translation $T : D \rightarrow Z$ and a scientific target $q : D \rightarrow Y$, there exists a function $g : T(D) \rightarrow Y$ satisfying $q = g \circ T$ on D if and only if $q(x) = q(x')$ whenever $x, x' \in D$ and $T(x) = T(x')$.*

Proof. If $q = g \circ T$, equality of translated outputs implies equality of targets. Conversely, when q is constant on the inverse image of every $z \in T(D)$, define $g(z) = q(x)$ for any x with $T(x) = z$. The stipulated constancy makes the definition independent of the choice of x , and hence $q = g \circ T$. $\square$

Proposition C.1 specializes the factorization result of Proposition A.3 to *translation between representations*. Its scientific significance is task-relative: the mean does factor through the translation in Equation (81), while the maximum generally does not. Adding precise metadata about the averaging method can reveal the insufficiency, but cannot recover the discarded contrast without further information. Similarly, an AI system learning a plausible distribution of underlying observations from training data would be estimating a conditional distribution, not performing guaranteed lossless reconstruction of the omitted samples. Validation should report that distinction rather than interpreting confident model outputs as recovered evidence.

C.3 A Quantitative Limit on Approximate Reconstruction

Exact factorization can fail even when a receiving system is permitted to return an approximation. A uniform error tolerance requires additional information about the target's variation within the records discarded by the translation. In the simplest finite scalar case this limit can be stated without statistical assumptions.

Proposition C.2 (Minimum uniform loss for a finite scalar target). *Let D be a nonempty finite set, $T : D \rightarrow Z$ a deterministic translation and $q : D \rightarrow \mathbb{R}$ a target. For each $z \in T(D)$, let*

$$\Delta_q(z) = \max_{x:T(x)=z} q(x) - \min_{x:T(x)=z} q(x). \quad (82)$$

There exists a reconstruction $g : T(D) \rightarrow \mathbb{R}$ satisfying $|q(x) - g(T(x))| \leq \varepsilon$ for every $x \in D$ if and only if $\Delta_q(z) \leq 2\varepsilon$ for every $z \in T(D)$, for $\varepsilon \geq 0$.

Proof. Fix a translated output z . If a proposed $g(z)$ is within ε of every source target in that fiber, its maximum and minimum target values each lie at distance at most ε from $g(z)$, so the difference is at most 2ε . Conversely, if the difference is at most 2ε , set $g(z)$ to the midpoint of the maximum and minimum target values. Every target value in the fiber then lies at distance at most ε from the midpoint. Repeat for all fibers. $\square$

For the mean-preserving example in Equation (81), restrict the candidate domain to $D = \{(0, 0), (-2, 2)\}$ and let q be the maximum. The translated output is zero in both cases, with target values zero and two. Hence the minimum achievable worst-case absolute error is one. A statistical prior or a trained AI reconstruction method may prefer one source, but it cannot provide a smaller *guaranteed* error on this stipulated domain without additional distinguishing information. On an infinite domain or a target valued in a general metric space, the finite scalar midpoint criterion is not sufficient as stated; the appropriate requirement instead concerns the existence of an admissible center for every translation fiber. The stated proposition deliberately makes no universal assertion about probabilistic recovery or legal interpretation.

C.4 Sampling Support and Reversal under Aggregation

A second error arises without incorrect arithmetic or loss of numerical precision. Consider two observation strata U and V , possibly representing two geographically distinct parts of a monitoring network. In a first represented survey, the positive-event rates are 0.90 in U and 0.10 in V . In another survey, the corresponding rates are 0.95 and 0.15. Each stratum's rate is greater in the second survey. However, if the first survey weights U at 0.9 and V at 0.1, while the second weights U at 0.1 and V at 0.9, the aggregates are

$$\bar{p}_1 = 0.9(0.90) + 0.1(0.10) = 0.82, \quad \bar{p}_2 = 0.1(0.95) + 0.9(0.15) = 0.23. \quad (83)$$

The aggregate declines while both stratum-specific rates rise. This is a standard aggregation-reversal mechanism illustrated here with fictional observational numbers; it does not imply an actual historical environmental trend. The separate surveys are indexed by a representational comparison relation, not by a metaphysically privileged universal time. A federation lacking sampling-weight metadata could portray the aggregate decline as contradicting the stratum-level increases even though all figures are internally consistent.

A legitimate comparison therefore requires explicit sampling frame, target population, aggregation weights, unit, coordinate convention, and support. Reweighting to a common reference population can sometimes answer a specified comparative question, although that operation changes the estimand and may require assumptions about representativeness. When those assumptions cannot be justified, the appropriate outcome is qualified noncomparability rather than silent translation. This example cautions against treating shared variable names or model-generated homogenization as sufficient evidence of cross-region equivalence.

C.5 Observational Agreement and Divergent Generative Models

A third limitation concerns models rather than the recorded variables themselves. Let a chosen coordinate u index locations or sampling positions along a transect. Two candidate functions are $f_A(u) = 0$ and

$f_B(u) = u(u - 1)$. They agree at the sampled positions $u = 0$ and $u = 1$, but differ at an unsampled position $u = \frac{1}{2}$, where $f_B(\frac{1}{2}) = -\frac{1}{4}$. Identical observed outputs at the selected supports therefore leave distinct possible continuations of the fitted representation. The example does not privilege u as physical time; the same mathematical issue occurs across spatial coordinates, experimental conditions, and relational observation supports.

Let $\mathcal{M}$ be a candidate class of models, S the selected observation operator, and Q a scientific query about model implications. Observational equivalence is the relation $m \sim_S m'$ when $S(m) = S(m')$. To infer Q uniquely from the available observations, it is necessary that Q be invariant within each $\sim_S$ class. This restates an identifiability requirement at the level of model representations and applies whether models are expressed by statistical maps, differential equations, simulation systems, or other generative procedures.

Proposition C.3 (Observational agreement does not determine an unobserved query). *If there exist $m_1, m_2 \in \mathcal{M}$ with $S(m_1) = S(m_2)$ but $Q(m_1) \neq Q(m_2)$, no rule using only $S(m)$ can return $Q(m)$ correctly for every $m \in \mathcal{M}$.*

Proof. Any such rule must assign the same output to identical observed values $S(m_1)$ and $S(m_2)$. Since the required target outputs differ, it must fail for at least one model. $\square$

Proposition C.3 is logically elementary, but has practical consequences for knowledge infrastructure. Increasing a semantic mapping’s apparent precision cannot resolve a structural lack of identifying observations. Likewise, consensus among several models fitted to the same selected records does not establish uniqueness of the scientific account unless the candidate model class and the identifying information are examined. A federation should preserve alternative models, declared priors or assumptions, validation support, and disagreements relevant to the intended query. Independent measurement or carefully targeted experiments may help separate representations that current data leave observationally indistinguishable, although feasibility and ethical authorization must be evaluated independently.

C.6 Composing Translations and Allocating Error Budgets

In a federated infrastructure, data may pass through several representations: a native sensor record, an institutional archive, a harmonized scientific dataset, and a model-ready exchange product. Suppose $T_{AB} : D_A \rightarrow X_B$ and $T_{BC} : D_B \rightarrow X_C$ are translations, and q_A, q_B, q_C produce comparable quantities in a metric space (Y, d) . Assume $T_{AB}(D_A) \subseteq D_B$. If, for a given $x \in D_A$, the two stages satisfy $d(q_A(x), q_B(T_{AB}(x))) \leq \varepsilon_{AB}(x)$ and $d(q_B(y), q_C(T_{BC}(y))) \leq \varepsilon_{BC}(y)$ at $y = T_{AB}(x)$, a triangle inequality yields a composed bound.

Proposition C.4 (Task-relative error bound for composed translations). *Under the stated domain and common-metric assumptions, the composite $T_{BC} \circ T_{AB}$ obeys*

$$d(q_A(x), q_C(T_{BC}(T_{AB}(x)))) \leq \varepsilon_{AB}(x) + \varepsilon_{BC}(T_{AB}(x)). \quad (84)$$

Proof. Insert the intermediate value $q_B(T_{AB}(x))$ and apply the triangle inequality in (Y, d) , followed by the assumed stagewise bounds. $\square$

The upper bound in Equation (84) is a statement about a *particular target* and a common comparison metric. It does not imply that a complete dataset, generative process, or governance meaning survives repeated conversion. It also ceases to apply if the first output is outside the validated domain of the second, or if q_A, q_B , and q_C refer to different scientific quantities. When translation introduces a model with uncertain parameters, an appropriate uncertainty propagation calculation requires further conditions (for instance a local stability or Lipschitz assumption); the additive bound alone does not prove robustness to such model variation.

As a deliberate failure case, consider a threshold interpretation $q(y) = \mathbf{1}\{y > 0\}$. Arbitrarily small perturbations of a numeric value near zero can change the classification completely. A small absolute

translation error in the underlying numeric measurement therefore need not imply a small error in this derived categorical decision. Stating a tolerance in temperature units cannot automatically certify adequacy for a threshold-based warning service. Validation should include the decision margin and uncertainty around the relevant boundary. Similarly, two services may map vocabularies in opposite directions without their round trip preserving scientific meaning: $T_{BA}(T_{AB}(x))$ can differ from x in ways that matter for the target query even if both exchanges are syntactically valid.

C.7 Concept Mappings, Context, and Institutional Categories

Consider two catalogues that both attach the label “coastal habitat” to observations. Catalogue A uses it for a physical class of surveyed environments, while catalogue B uses a legal or management category with access restrictions and designated administrative boundaries. A terminological mapping may support discovery of relevant records, but it cannot legitimately import the legal definition, conservation status, or permitted-use rule into the other system. These distinctions are not measurement noise. They concern different kinds of predicates, different evidentiary practices, and different competent authorities.

SKOS explicitly permits distinct mapping relations rather than requiring every cross-vocabulary match to be treated as identity [70]. In a proposed federation, a vocabulary link should record its declared direction, mapping type, domain of intended use, version, asserting agent, and contestation status. A particularly consequential error occurs when a *broaderMatch* or merely related concept is silently interpreted as a logical equivalence, allowing a downstream application to substitute one category for another. Semantic conformance cannot adjudicate such an institutional leap: a legitimately governed catalogue must make the mapping’s author and justifications challengeable.

The problem becomes more acute for heterogeneous generative representations. A trained classifier, a mechanistic environmental simulator, and a community-conducted descriptive inquiry may produce apparently corresponding terms while relying on different methods for recognizing a phenomenon. Their outputs may be useful together under a task-specific protocol, but a one-to-one term dictionary does not provide a generally valid translation between the systems’ inferential operations. It is therefore appropriate for a federation to publish a *translation warrant*, which identifies the tested tasks, excluded cases, validation evidence, responsible reviewers, and routes for revision, rather than to certify universal interoperability.

C.8 Sensitive Derivatives and Auxiliary-Information Leakage

A technically valid and accurate translation can itself introduce an unacceptable disclosure. Suppose a released aggregate contains the combined number of observations from two sensitive locations. If one location’s contribution is already public, a recipient may infer the other by subtraction. Merely suppressing the name or coordinates of the second location does not necessarily prevent reconstruction. Similarly, a model output may reveal a protected ecological feature indirectly, even when none of the input records is redistributed openly.

Let D denote a restricted underlying dataset, $T(D)$ a candidate public output, a the recipient’s auxiliary information, and $s(D)$ a sensitive target. A sufficient condition for exact inferential exposure is the existence of a reconstruction rule h on the relevant domain such that

$$s(D) = h(T(D), a). \quad (85)$$

Equation (85) is not a definition of privacy. When no such deterministic h exists, substantial probabilistic disclosure can remain possible. Conversely, claiming that some recipient can infer s requires specification of their available information and inference assumptions. In the elementary example, $T(D) = n_1 + n_2$ and auxiliary information $a = n_1$ permit recovery $n_2 = T(D) - a$ exactly. All quantities are fictional and do not contain real ecological locations.

Rights and disclosure tests therefore apply to *derived artifacts*, their intended uses, and their audience context, rather than being inherited automatically from source-record policy. The same principle

applies to legitimate scientific accountability: a valid provenance chain does not compel a steward to disclose a protected input. Well-specified restricted review, safe aggregation, and contestable decisions can sometimes reconcile reproducibility with rights, but no single technical restriction is adequate for all scientific and governance purposes. This continues the source-versus-derivative distinction of Appendix B and Section 9.

C.9 Cross-Register Agreement and Manifestational Underdetermination

The three coupled reality registers used in Section 2.4 help specify what a translation test can establish. Suppose systems A and B have manifest records m_A, m_B and task-dependent observation or production maps π_A, π_B from their selected relational configurations r_A, r_B . A translation may preserve the selected output q , even when the generating configurations remain distinct:

$$q_A(m_A) = q_B(T_{AB}^{(q)}(m_A)) = q_B(m_B), \quad m_A = \pi_A(r_A), \quad m_B = \pi_B(r_B). \quad (86)$$

The equality in Equation (86) is an explicitly selected cross-system invariant. In general, it does not entail $r_A = r_B$ or an isomorphism of the systems' full relational histories. To see why, consider a sensor network encoding a binary alert as a numeral and another permitted communication process encoding the same alert through a repeatable chemical marker. Under a stipulated reader both can recover the alert bit, while their generating mechanisms, materials, failure modes, and possible meanings of the signal are different. The example is a logical construction, not an empirical report about organisms or extraterrestrial cognition. A stronger equivalence conclusion would require independently justified bridges among practices, histories, and generative mechanisms.

Manifestational difference likewise fails to prove distinct underlying ontologies. Different mathematical presentations may be definitionally equivalent or preserve selected structures, yet neither a verified mapping nor shared predictive success alone settles ultimate existence. Ontological questions remain bracketed unless the inquiry supplies and defends a stronger bridge principle. In applied federation design, validation should therefore distinguish preservation of a selected scientific quantity, preservation of the inferential operations used to evaluate it, recognition of the institutionally authorized practice, and the much stronger claim that two systems have the same generative language. These criteria can be logged and contested separately.

C.10 Counterexample Register and Evaluation Protocol

Table 28 summarizes the deliberately bounded failures. Each case includes a potentially testable observation, rather than using an abstract claim of "semantic difficulty" to evade falsification.

Table 28: Hypothetical translation counterexamples and task-relative validation requirements.

Failure mechanism	Counterexample	Required test or qualification
Aggregation loss	Mean preserved, maximum not recoverable	Compare target across records in the same translation fiber
Support shift	Within-stratum rates rise while weighted aggregate falls	Preserve weights, sampling frame, and estimand
Model nonidentification	Models agree on selected observations, diverge elsewhere	Test observability of the target and alternatives
Error amplification	Small numeric error flips a threshold category	Inspect margin, stability, and intended decisions
Concept mismatch	Scientific label mapped onto a legal category	Validate mapping type, use scope, and authority
Disclosure through derivatives	Public sum plus external input reveals a protected term	Evaluate output and auxiliary information together

A practical cross-node evaluation should assemble a versioned test corpus and a specification of tasks before measuring interoperability. For each translation, the evaluation record should state its native input types, calibrated measurement support, intended comparison quantity, validated domain, observed loss, authorization context, and correction procedure. Negative tests should be mandatory: identical means with distinct extremes, incompatible spatial supports, changed survey weights, drifted sensor calibrations, divergent models, unauthorized concept promotion, and sensitive outputs paired with plausible auxiliary information. Independent reviewers should be able to dispute not only the factual results but also the selection of the validation tasks and affected communities.

A small executable test suite accompanying the source archive checks only the finite arithmetic counterexamples, model nonidentification example, and triangle-inequality composition example. Such passing checks would confirm that the illustrative computations have been transcribed correctly; they could not establish whether a global epistemic translation mechanism is valid in practice. Empirical research would require real observational programmes, metrological validation, carefully documented permissions, domain experts, and participatory review where communities are affected. Even a successful limited deployment would not imply that heterogeneous sciences share a single generative grammar. Our claim is more circumscribed: a federated infrastructure can make the *conditions and limits* of attempted translations explicit, testable, revisable, and institutionally contestable.

D Illustrative International Observation Networks

This appendix examines small, deliberately fictional observation networks whose components resemble the *roles* of meteorological stations, oceanographic platforms, satellite receivers, archives, metadata catalogues, caches, analytical facilities, and independent reviewers. Its purpose is to separate four properties that are frequently conflated in discussions of global scientific infrastructure: physical or digital connectivity; actual discovery and delivery of records; scientific adequacy for a specified question; and institutionally authorized use, correction, and action. The examples are not a reconstruction of any country's observation infrastructure, not a deployment proposal, and not measurements of the resilience of WMO, Argo, GOOS, or other named programmes. Their finite graphs test only the assumptions declared here. The distinction matters because established initiatives demonstrate particular exchange and coordination mechanisms without supplying a universal guarantee of epistemic or institutional interoperability [65, 64, 39].

The appendix builds on the observation event and role vocabulary of Appendix A, the record profile of Appendix B, and the translation counterexamples of Appendix C. The federation described in Section 5 provides the architectural context. We first specify typed network layers and access rules, then present an illustrative nine-role federation, analyse several failures, and describe a reproducible finite test. *Scientific events are not identified with network nodes*: the graph's vertices denote configured service roles, while particular measurements, reviews, and authorizations remain events that can be attributed to those roles and revised in a relational history.

D.1 Service Roles, Observation Events, and Typed Links

Let the finite set of service roles be

$$N = \{L, M, T, A, B, C, K, P, V\}, \quad (87)$$

where L is a fictional land station, M a moored marine observing platform, T an orbital-observation reception service, A and B distinct regional archives, C a discovery catalogue, K a data cache, P a scientific analysis service, and V a review service. These symbols abbreviate roles in a finite service configuration, not fixed or enduring ontological entities. One organization may operate multiple roles; the personnel, equipment, and organizational authority associated with a role may change while the assigned role identifier remains the same for the purpose of one test. Conversely, the same represented organization may participate in several distinct events.

The model records three typed directed graphs on N : a data-delivery graph $G_D = (N, E_D)$, a discovery graph $G_F = (N, E_F)$, and an independent-review graph $G_R = (N, E_R)$. A fourth relation $\text{Perm}(s, r, u, \gamma)$ represents whether a governed record version r from source role s may be used by role u for purpose γ under a specified policy interpretation. Separate metadata describe measurement suitability, version status, and decision authority. The graphs need not have the same edges or share their failure modes.

$$\mathcal{N} = (N, G_D, G_F, G_R, \text{Perm}, \mathcal{H}, \Theta), \quad (88)$$

where $\mathcal{H}$ refers to the typed relational history of Appendix A and Θ is a dossier of quality conditions, service dependencies, failure domains, and the declared scientific task. Equation (88) treats access policy as a contextual relation rather than pretending it follows automatically from a digital edge. Scientific relevance must be evaluated against Θ : two records can be delivered successfully and still have incompatible spatial support, units, calibration history, or intended use.

For a required record r , define the route qualification predicate

$$\text{Use}(r, u; \gamma) = \text{Found}(r, u) \wedge \text{Delivered}(r, u) \wedge \text{Perm}(s(r), r, u, \gamma) \wedge \text{Fit}(r, \gamma), \quad (89)$$

where the component predicates are evaluated in a stated network configuration. Fit is not a universal truth predicate: it stands for task-specific evidence about provenance, uncertainty, calibration, measurement support, and the suitability of the record for a particular inference. In Equation (89), Found means discovery through currently available metadata services, or through a previously verified and still-valid direct endpoint. A record may be physically retrievable without satisfying the discovery term for a first-time user. Similarly, Delivered may be true even when institutional authorization is absent.

D.2 A Small Coastal, Atmospheric, and Orbital Federation

In the hypothetical baseline, observations from L and M are deposited at A , while T deposits observations at B . Both archives publish discovery metadata to C , deliver permitted data to a cache K , and offer direct data paths to the analysis service P . The cache can deliver previously synchronized versions to P . The review service V receives scientific claims from P and can retrieve source versions separately from A and B in order to check particular claims. Figure 4 distinguishes selected data, discovery, and review edges by line style. Permissions and scientific-compatibility conditions are deliberately omitted from the drawing and specified in prose, since drawing those conditions as additional undifferentiated edges would invite the false interpretation that delivery and authority are the same relation.

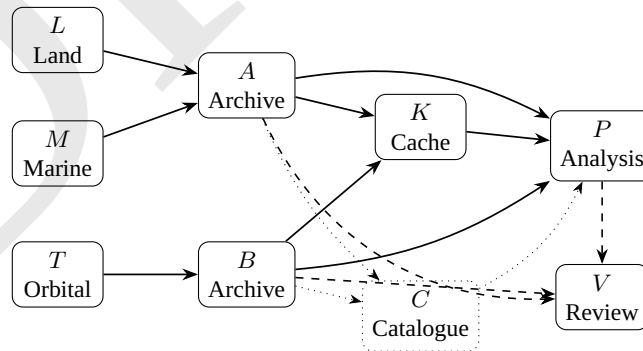


Figure 4: Fictional international observation-service network. Solid edges represent data routes, dashed edges illustrate selected review dependencies, and dotted edges show discovery notifications. The figure omits authorization and scientific-compatibility predicates, which require separate evaluation.

The fixed example contains three source families $S = \{L, M, T\}$. Its data and review edges are

declared, respectively, as

$$\begin{aligned} E_D &= \{LA, MA, TB, AK, BK, KP, AP, BP\}, \\ E_F &= \{AC, BC, CP\}, \\ E_R &= \{AV, BV, PV\}, \end{aligned} \tag{90}$$

where juxtaposition XY denotes the directed edge $X \rightarrow Y$. All nodes are initially available and all three source versions are stipulated to have a valid baseline permission and quality status for a single hypothetical regional-analysis question. These stipulations do *not* claim that atmospheric, marine, and orbital records are scientifically interchangeable. A real comparison would require the sampling and translation tests of Sections 4 and 6. Under the baseline stipulations, the data paths $L \rightarrow A \rightarrow P$, $M \rightarrow A \rightarrow P$, and $T \rightarrow B \rightarrow P$ make all three represented source families deliverable to P .

D.3 Connectivity, Discovery, and Authorized Scientific Use

For a directed typed graph G , let $s \rightsquigarrow_G t$ denote existence of a path under a declared set of available vertices and edges. Raw connectivity is the indicator $R_D(s) = \mathbf{1}\{s \rightsquigarrow_{G_D} P\}$. Fresh discoverability of a source s requires its archive's metadata path to C and the catalogue path to P ; the predicate can be different for an existing consumer retaining a verified endpoint. For the first-time consumer used in the simulation, define

$$R = \sum_{s \in S} R_D(s), \quad F = \sum_{s \in S} \mathbf{1}\{a(s) \rightsquigarrow_{G_F} P\}, \quad U = \sum_{s \in S} \mathbf{1}\{R_D(s) \wedge \text{Found}(s, P) \wedge \text{Perm}_s \wedge \text{Fit}_s\}, \tag{91}$$

where $a(L) = a(M) = A$ and $a(T) = B$. Equation (91) is an experimental protocol, not a scalar measure of epistemic credibility. In particular, an independently reviewable scientific output requires additional evidence about the review service V , its access to relevant source versions, and its freedom and competence to challenge P 's inference. We report that property separately instead of folding it into U .

Proposition D.1 (Authorized usability cannot exceed reachable delivery). *For any fixed finite set of source records, if the admissibility criterion for Use includes a successful data-delivery path as a necessary condition, then the number of usable records is at most the number of deliverable records. The converse need not hold.*

Proof. Every record counted as usable satisfies the delivery predicate in Equation (89). Hence the usable set is a subset of the deliverable set. For the converse, choose a deliverable record whose purpose-specific permission is false, or whose calibration evidence is inadequate. The physical data route remains intact while Use is false. $\square$

Proposition D.1 is elementary but is precisely the inference that generic network diagrams often obscure. With identical cables and endpoints, changing an authorization decision can change usable scientific access from one record to zero. Conversely, permission may be granted while a damaged sensor or non-operating archive prevents delivery. An institution with strong network access but weak computational capacity may also fail to reproduce the published inference, even after all raw records are delivered. Section 9 examines the relevant differences between record access and effective epistemic participation.

D.4 Discovery Cut Sets, Caches, and Prior Knowledge

The service C is the only discovery catalogue in the fixed baseline. Removing it disconnects the archives from a *new consumer's discovery process*, but it leaves the direct data paths $A \rightarrow P$ and $B \rightarrow P$ intact. This case demonstrates why operational data availability and newly obtainable knowledge of a resource's

address can diverge. It also demonstrates the danger of making a catalogue a hidden gatekeeper: describing the whole architecture as “distributed” does not ensure that users have alternative ways to discover resources.

Proposition D.2 (Single-catalogue discovery cut). *Suppose every currently admissible first-time discovery path from a publisher set Q to a consumer p passes through a catalogue vertex c , and no independent previously trusted endpoint list is available to that consumer. If c becomes unavailable, no record from Q can be newly discovered under that protocol, even if a data-delivery route from a publisher to p survives.*

Proof. By hypothesis, every first-time discovery path contains c . Removing c removes all such paths. The conclusion about data delivery is independent because delivery paths belong to another graph and need not pass through c . $\square$

Proposition D.2 is a conditional graph-cut statement, not a claim that actual WIS2 has a unique un-replicated catalogue. WIS2 specifies publisher nodes and global services that include catalogues, brokers, caches, and monitoring; its operational redundancy must be assessed from the system’s real deployments and procedures [64, 65]. An alternative trusted catalogue, independently verified manifest, or known direct endpoint would defeat the premise of Proposition D.2. A cache likewise improves resilience only for particular records and versions already held and valid under the relevant release policy. A public cache does not imply access to restricted data, guaranteed freshness, or authority to reuse data after a policy revision.

D.5 Redundancy, Common-Cause Failures, and Dependency Domains

The delivery graph has a short alternative path for L and M : their archive can reach the analysis service both directly and through K . Removing only the direct link $A \rightarrow P$ therefore leaves $A \rightarrow K \rightarrow P$ available. Removing the cache alone leaves the direct links available. The same graph can, however, contain a highly consequential archive dependency: both land and marine sources rely on A . Thus an outage of A simultaneously affects two otherwise distinct observing programmes. The same mechanism can appear across domains if nominally separate servers share a physical energy supply, cooling infrastructure, cloud region, or credential authority.

Proposition D.3 (Disjoint service paths and a single service failure). *Let s, t be available endpoints of a directed service graph. If there are two s – t paths with no shared internal vertices, deletion of any one internal vertex leaves at least one of these two paths intact. This conclusion need not survive a failure that removes multiple vertices from a shared physical or institutional dependency domain.*

Proof. A vertex internal to at most one of the two internally vertex-disjoint paths cannot belong to the other path. Removing it therefore leaves the other path. For a shared-domain failure, select one internal vertex from each path and suppose both rely on a common power supply. Removing the two vertices at once defeats both routes, despite their graph-theoretic disjointness. $\square$

The conclusion of Proposition D.3 is deliberately narrower than a reliability guarantee. Different routes may share the same archive, software implementation, governmental access restriction, or un-modelled physical dependency. A realistic failure analysis therefore needs an additional hypergraph or incidence relation $\text{DependsOn} \subseteq N \times \mathcal{F}$ mapping service roles to dependency domains $\mathcal{F}$, together with an assessment of correlated events. Assuming independent edge failures would make redundancy seem stronger than warranted when the dependency structure is unknown. A different source of resilience may arise from institutional capabilities: another scientifically qualified organization may be able to review or reconstruct observations even when it cannot reproduce the original platform’s exact measurement events.

D.6 Finite Stress Tests and Interpretable Outcomes

Table 29 reports a set of deterministic tests on Equation (90). The table counts available *source families*, not the number or quality of individual observation records. R counts families with a data route to P ; F counts families discoverable by a first-time consumer using the current catalogue; U counts families satisfying all baseline delivery, discovery, permission, and fitness predicates. For each scenario, service outages remove the affected vertex and all incident edges. A denied marine permission sets the use predicate for M to false without changing its data links. The last scenario removes the review-edge $B \rightarrow V$ to illustrate loss of independent source access for the orbital family even when the three displayed counts remain unchanged.

Table 29: Deterministic outcomes on the hypothetical nine-role network; all counts range from zero to three.

Scenario and intervention	R	F	U	Interpretation and uncounted limitation
Baseline, full permissions	3	3	3	Scientific validity still requires external assessment
Catalogue C unavailable	3	0	0	Known endpoint access may survive; first-time discovery fails
Archive A unavailable	1	1	1	Land and marine source families share an archive dependency
Direct data link $A \rightarrow P$ unavailable	3	3	3	Cached route maintains reachability under stipulated cache state
Cache K unavailable	3	3	3	Direct archive paths preserve delivery
Marine-use permission denied	3	3	2	Delivery is not authorized use
Review edge $B \rightarrow V$ unavailable	3	3	3	Independent review of orbital source lacks stipulated direct access
Analysis service P unavailable	0	0	0	Observation acquisition may continue elsewhere

In the final row, discoverability is measured for consumer P , so the count becomes zero when that consumer is unavailable. Other consumers might continue operating. In the catalogue-outage row, $U = 0$ applies to the specified first-time workflow; it would not necessarily hold for a prior user who retains a still-authorized endpoint. In the cache and direct-link rows, the assumed route through K is valid only where a current and permitted copy has been synchronized. The graph cannot establish that such a copy actually exists, that latency is adequate for warning operations, or that a scientific product meets its uncertainty requirement. The last observation in Table 29 is likewise not a conclusion that an external reviewer loses every possible verification route: it concerns the one independent link included in the stipulated review layer.

A second test varies scientific rather than network conditions. Suppose the land station and the marine platform both publish compatible numerical fields but the marine series undergoes undetected calibration drift. Every service path, discovery notification, and permission in Equation (90) can remain intact. Yet an inference mixing these signals may be biased, and the network’s counts remain $R = F = U = 3$ only if the *assumed* fitness flag has not been updated. Independent calibration review could subsequently change Fit_M and reduce U to two. The episode demonstrates a limitation of a purely digital topology: scientific credibility depends on evidence and history, not on a graph’s edge count.

D.7 Cross-Border Rights, Accountability, and Effective Participation

Even the fully connected baseline leaves institutional questions unresolved. The operators of A and B may hold different responsibilities for public records, embargoed measurements, or community-governed ecological information. An analysis service may have permission for climate modelling but lack authorization to publish fine-grained species locations. A reviewer might be allowed to inspect protected data within a secure environment without receiving an unrestricted copy. The same record can have different permitted uses and versions as an institution’s rules, the applicable jurisdiction, or community decisions change. The record profile of Appendix B can preserve these claims and their provenance; it cannot establish their legal validity or normative legitimacy.

Similarly, effective participation depends on computation, staff capacity, language, methodological knowledge, and the ability to contest findings. Suppose archive operators provide identical access cre-

dentials to two universities, while one can run the relevant ocean model and the other lacks the required computation and calibration expertise. Their *formal* access may be equal while their ability to reproduce or challenge inferences is unequal. The ethical and governance criteria of Section 8 and Section 9 therefore complement the graph model. A credible federation needs mechanisms for alternative review pathways, transparent decisions, correction of metadata and models, responsible incident notification, and access to meaningful remedy; establishing these mechanisms remains an institutional project requiring empirical evaluation.

A useful experimental audit should report stage-specific outcomes rather than publishing a single “resilience” score. For each task, report observation acquisition and quality, new discovery, authorized delivery, compute availability, independent scientific reproducibility, institutional responsiveness, and ecological or community effects. It should also identify which actor is accountable when an assumption fails and how evidence about that failure enters the revision history. The fictional tests here show only that some of these dimensions can move independently in an explicitly stated network.

D.8 Scope of the Representation and Empirical Continuations

The appendix does not estimate real-world global observation coverage, outage rates, cross-border permissions, climate-prediction quality, or community trust. The graph has few nodes, stipulated capabilities, and deterministic interventions. It omits transmission latency, changing traffic loads, packet loss, measurement uncertainty distributions, organizational budgets, and adaptive behaviour. The conclusions of Propositions D.1, D.2, and D.3 follow under the stated predicates, but their applicability to an actual programme would depend on faithfully mapping its technical, legal, and scientific conditions. WMO’s distributed WIS2 services and the internationally coordinated observing networks tracked by OceanOPS illustrate useful distinctions between operating nodes, shared exchange services, operational metadata, and the activities of national contributors; their actual deployment details must be examined on their own terms [64, 39, 62].

The accompanying executable example checks the graph counts and the key counterexamples of Table 29. A subsequent empirical study should replace its fictional graph with authorized observations of system operation, identify independent failure domains, document scientific task and permission classes, and undertake controlled tests where they do not threaten safety-critical warning services. Institutional participants and affected communities should be able to challenge the selected tasks, access rules, and assumptions about acceptable failures. The primary methodological outcome is a distinction: globally connected observation services, scientifically credible knowledge, and legitimate cooperation are overlapping but separable accomplishments. A meaningful infrastructure theory should specify which one has been demonstrated in each case.

E Governance Scenarios and Dispute Procedures

This appendix tests a limited question: what can a federated observation infrastructure actually establish about a disputed publication, access restriction, or scientific correction when different participating institutions possess different claims to authority? It extends Section 8, the protected-derivative analysis of Section 9, the metadata profile in Appendix B, and the typed service dependencies of Appendix D. All organizations, communities, records, rules, and outcomes in the worked examples below are fictional. The scenarios are procedural thought experiments, not an account of the law of any jurisdiction, a proposal to override Indigenous or community authority, or evidence that one arrangement is universally superior.

The appendix has two objectives. One is analytical: to keep technical authentication, scientific merit, legally recognized authority, and ethically defensible governance separate. The other is constructive: to specify a minimal dispute protocol that can record rival positions, preserve evidence, request a provisional response, identify the competent decision forum, and report the effectiveness and limitations of an attempted remedy. A data catalogue may carry records of such steps, but recording a step does not

make the step just, authoritative, or executable. The distinction is especially important where sensitive ecological information, unequal computing resources, and potentially urgent public warnings meet [57, 6, 52].

E.1 Governance Units, Standing, and the Scope of a Claim

Assume a finite set of institutional role-bearers I , observation versions R , derived artefacts Y , scientific tasks Γ , and rule systems J . The role-bearers can include a record custodian, an instrument operator, a scientific author, a community or collective with recognized interests, a publishing service, an independent reviewer, and a remedial decision forum. Their role assignments may change with a relational history h ; they are not identified with event vertices in the underlying observation complex. Standing to file a dispute, competence to decide it, practical capacity to participate, and jurisdiction to impose a remedy need not coincide. Neither the mere possession of a file nor the ability to authenticate its provenance establishes any of those capacities.

Let a claim record be

$$c = (\iota, \alpha, \omega, \gamma, \chi, \sigma, \kappa, \epsilon, h_c), \quad (92)$$

where ι is a stable record identifier, $\alpha \in I$ is the claimant, $\omega \in R \cup Y$ is the contested object, γ specifies the intended use, χ is the requested change (for example restrict, correct, review, or restore), σ identifies the affected persons, organizations, or ecological interests insofar as they can be responsibly specified, κ records asserted authority and its scope, ϵ names the evidence dossier, and h_c links the filing event to its provenance history. Equation (92) represents an *assertion*; the assertions in κ and ϵ may be incomplete, incompatible, or disputed. Claims about nonhuman ecological effects may require a representative procedure rather than a fictional direct signature by a species or ecosystem.

A separately typed decision record may contain issuer, mandate, considered evidence, rule-version references, affected object versions, decision, provisional measures, reasons, review path, and verification-of-execution status. It is important to record the *scope* of a decision: an institution competent to correct its own sensor record may lack authority to compel another organization's model to be retrained or an independently operated service to retract a derivative. Rights of access, rights to publish, duties to correct, and power to bind third parties therefore require separate predicates. FAIR-oriented reuse and CARE-oriented collective authority illuminate different, potentially interacting concerns and should not be collapsed into an undifferentiated "open data" flag [61, 6].

E.2 Four Illustrative Institutional Arrangements

Hold fixed a small set of scientific data services and an identical dispute over a derived marine record. Vary the institutional procedure instead of the technical data graph. The following four arrangements are stipulated alternatives. *Repository-local governance* assigns initial correction and release decisions to each record custodian and exposes a documented escalation route. *Bilateral reciprocal governance* makes a data-sharing agreement between two organizations the specified source of review and remedy, while leaving outsiders' standing unresolved. *Federated multi-party review* uses a jointly constituted forum for defined cross-institutional classes of disputes, while primary scientific records and operational custody remain distributed. *Fragmented governance* has independent local rules and no accepted common dispute forum for some cross-border conflicts. No arrangement has an automatic ethical ranking; each may differ in jurisdictional enforceability, affected-community participation, emergency responsiveness, resources, and exit conditions.

The comparison in Table 30 concerns *authority topology*: who can ask, decide, review, and execute. It is different from the packet-routing and discovery topology analyzed in Appendix D. Under one topology, access to a record may continue while no institution accepts competence to adjudicate a dispute about its use. Under another, a shared appeal forum may exist on paper while a community lacks translation assistance, funds, secure computational access, or a means of presenting evidence. Polycentricity describes the possibility of multiple centers of rule-making and coordination; it does not establish

Table 30: Hypothetical governance arrangements under the same observation network and disputed derivative.

Arrangement	Initial decision locus	Potential route for cross-party dispute	Distinctive limitation to evaluate
Repository-local	Originating custodian	Escalation under custodian’s published procedure	Custodian may not bind a downstream publisher
Bilateral reciprocal	Joint agreement parties	Designated contractual review forum	Affected outsiders may lack standing or remedy
Federated multi-party	Mandated shared review body	Defined appeal and correction workflow across members	Participation cost, mandate boundaries, and enforcement
Fragmented	Separate local authorities	Negotiation if all parties agree	Conflicting permissions can persist without resolution

that their decisions are mutually consistent or that weaker participants automatically receive effective remedies [51, 52].

E.3 Conflict of Rules and Decision Authority

Consider a data-use request $q = (\omega, \gamma, u, h)$ concerning a versioned object, purpose, user, and relevant history. For institution i , let $p_i(q)$ belong to {permit, deny, undetermined}. The vector of available policy positions is

$$\mathbf{p}(q) = (p_i(q))_{i \in I_q}, \quad \text{Resolve}_j : \mathbf{p}(q) \mapsto \{\text{permit, deny, pending}\}, \quad (93)$$

where I_q lists the institutions with potentially applicable authority and j indexes an explicitly recognized conflict-resolution rule. The rule cannot be derived from metadata syntax. A priority rule, consent requirement, appeal, emergency exception, or abstention procedure must be justified by an applicable governance arrangement and must state whose authority it claims to recognize. Equation (93) is deliberately not a claim about the lawfulness of any particular priority order.

Proposition E.1 (Policy records do not determine a unique joint authorization). *Suppose a request has two applicable, authenticated policy records with contradictory outcomes, one permitting and one denying the same purpose-specific release, and no mutually accepted rule that resolves their priority. Then the records alone do not determine a unique federation-wide authorization outcome.*

Proof. Consider the same authenticated pair (permit, deny). A hypothetical resolution procedure giving priority to the permitting decision outputs permit, whereas a procedure requiring agreement of all competent parties outputs deny or pending. Both procedures receive identical policy records. Without a specified and recognized selection rule, the observed metadata do not distinguish the outputs. The proposition concerns logical underdetermination, not whether either candidate procedure would be legally or ethically acceptable. $\square$

Proposition E.1 has a practical design implication: a federation’s API should be able to represent a *pending conflict* without quietly converting it into a permission. A safe interim response for an unresolved sensitive release may be restricted processing or postponement, but even that choice is a provisional measure requiring institutional justification, prompt review, and consideration of the harms of delayed scientific action. An emergency publication procedure must similarly declare its legal and procedural basis, the minimum information released, its expiry or review condition, and subsequent avenues of contestation. The point is to make the tradeoff accountable, not to install a technical veto or emergency override by default.

E.4 A Dispute Protocol with Reviewable Transitions

For a specified governance arrangement, define a dispute dossier state s in the finite set $\mathcal{S}$ whose possible labels are filed, notified, preserved, interim, reviewed, decided, remedy attempted, appealed, and closed. These labels denote roles of events in a workflow, not stages of a universal social time. A particular procedure may allow concurrent evidence preservation and notification, appeals after an initial decision, or a reopened matter after new evidence. The transition relation $T_j \subseteq \mathcal{S} \times \mathcal{E} \times \mathcal{S}$ must be indexed by the applicable procedural rules and accompanied by a ledger of evidence and actors; an unqualified global ordering of events is unnecessary.

An illustrative state transition may be permitted only if its role-specific guard is satisfied:

$$(s, e, s') \in T_j \implies \text{Guard}_j(e, s, \text{mandate}, \text{evidence}, \text{notice}) = 1. \quad (94)$$

For example, *decided* should normally require identification of the decision issuer, the relevant rule version, a reasoned determination, evidence considered, any refused evidence with explanation, and a review route. *remedy attempted* is not the same as *remedy executed*; execution needs separate evidence from the services and institutions affected. A restrictive provisional action must disclose its stated purpose, proportionality rationale, and a procedure for reconsideration, subject to justified confidentiality constraints. Equation (94) is a proposed software-validation condition, not proof that the resulting procedure is fair.

The state model intentionally admits *unknown*, *disputed*, and *not applicable* values where appropriate. It does not silently fill missing consent or standing data with permission. Nor should the state labels imply that disputes are necessarily resolved in a linear sequence: the relational history may contain branching reviews, parallel scientific assessments, new submissions, and superseded decisions. A machine-readable dispute record should preserve those relationships and the accessibility of the reasons to affected participants. Metadata specifications such as W3C PROV help represent derivation and responsibility assertions, but their structural validity remains distinct from the authority and effectiveness of an institutional decision [71, 74].

E.5 Scientific Corrections, Retractions, and Effective Remedies

Disputes about knowledge can concern the reliability of a sensor, translation of a variable, release of a sensitive derivative, authorship or contributor recognition, misuse of community-governed data, or the validity of an AI-generated inference. These disputes need different remedies. Correcting an observation value requires a quality-control and versioning process. Retracting a public derivative may require notifying downstream consumers, updating model registries, and identifying outputs already produced. Recognition and participation disputes may require changes to attribution, review procedures, or access to analysis infrastructure. Ecological harms may require measures outside the information system entirely. A single generic “resolved” bit cannot encode these distinct obligations.

Let $\mathcal{A}(c)$ be the set of remedial actions ordered for claim c and let $\mathcal{X}_h(a)$ be the available execution evidence for each action a at history h . A minimal representation of completion is

$$\text{Completed}(c, h) = 1 \implies \forall a \in \mathcal{A}(c), \text{VerifiedExecution}(a, \mathcal{X}_h(a)) = 1. \quad (95)$$

The reverse implication need not hold: executing the enumerated actions may leave unrecognized affected parties, irreversible dissemination, or harms outside the decision body’s mandate. Equation (95) is therefore a conservative reporting condition on the specific ordered remedies, not a definition of substantive justice.

Proposition E.2 (Complete technical traceability does not entail effective remedy). *There are two institutional arrangements with identical authentic provenance records for an erroneous released derivative and the same recorded correction decision, but different levels of effective remedial execution.*

Proof. Fix the record versions, their provenance DAG, the discovery metadata, and a decision requiring the publisher to correct a derivative and notify recipients. In arrangement A , the publisher can update the

affected service and reach the relevant recipients; verification evidence confirms both ordered actions. In arrangement B , the publisher lacks authority over the service and cannot identify a downstream redistributor. The same provenance and decision records exist in each arrangement, but the ordered actions are verified only in A . Therefore provenance and a recorded decision alone do not determine effective remedy. $\square$

Proposition E.2 complements the main-text result in Section 8. Technical integrity may make facts contestable, but does not supply legal standing, translation assistance, practical access to a tribunal, enforcement capability, or reversal of a disclosure. A federation should measure these capabilities instead of treating correction metadata as their substitute. In particular, an independently operated downstream model may encode sensitive source information even after the source record is restricted; removal of the source from public discovery does not mechanically erase the derivative, as Section 9 discusses.

E.6 A Worked Hypothetical Coastal-Observation Dispute

Consider a fictional marine observatory M publishing a nearshore temperature record r_1 through archive A . An analysis laboratory P uses r_1 and a coastal survey record r_2 to publish a geographically detailed risk map y_1 . A coastal stewardship body C subsequently claims that y_1 exposes a sensitive ecological location and that the declared data-use purpose did not authorize this level of disclosure. Separately, a monitoring scientist S detects a possible calibration drift in r_1 , which could affect the magnitude of a public heat-risk estimate. These are *two different claims*: one concerns authorization and ecological incidence; the other concerns scientific validity. Neither automatically decides the other.

Assume a shared review rule that grants C standing to challenge disclosure and S standing to request measurement review, without presuming that their assertions are correct. The review body can impose a temporary restricted-view mode on the federated portal only if its mandate covers that service. It can require the observatory to preserve r_1 and relevant calibration evidence, ask the laboratory for a reproducible derivation record, and provide the parties with reasons and an accessible opportunity to respond. After scientific review, the observatory publishes a corrected version r'_1 if justified, and the laboratory assesses whether y_1 must be recalculated. After an independent disclosure review, the body may conclude that a less granular derivative y_2 is appropriate for a particular public-warning purpose while protected detail remains available only in a controlled research setting, again subject to the applicable rules.

Table 31: Fictional coastal dispute: a bounded procedural trace and the evidence it cannot itself supply.

Recorded step	Required evidence in the hypothetical dossier	Remaining question
Filing and standing	Claims by C and S ; applicable standing rules; affected versions	Are the standing rules legitimate and inclusive?
Evidence preservation	r_1, r_2, y_1 identifiers, calibration and derivation manifests	Are the scientific records complete and reliable?
Provisional response	Competent issuer, narrow release restriction, review conditions	Is the restriction justified and proportionate?
Scientific decision	Drift assessment and published r'_1 if warranted	Does the correction support the revised inference?
Disclosure decision	Purpose, ecological sensitivity, reasons for chosen precision	Are affected interests adequately represented?
Remedy verification	Portal update, downstream notices, controlled derivative audit	Which independently circulated copies persist?

The trace in Table 31 illustrates why provenance and dispute adjudication require linked but different records. The calibration issue might be resolved with no change to access policy, while the disclosure issue may survive even after calibration is corrected. An emergency warning could justify public release of some coarsened information under a specific rule, but the exception would need an articulated basis and subsequent review. It cannot be inferred from the mere existence of a risk map. A review body

unable to reach downstream recipients should record its incomplete execution rather than claiming that the disclosure has been reversed. The result must remain provisional where the pertinent institutional competencies or affected interests have not been established.

E.7 Participation Capacity, Appeals, and Institutional Exit

The existence of an appeal form does not establish that a participant can actually use it. A fictional claimant may need interpretation into a working language, staff time, access to the restricted evidence in a secure setting, resources to reproduce model outputs, and protection against retaliation or exclusion. These are practical conditions for contestability. Let G_Q be the directed graph of notice, representation, submission, review, and appeal services available to a particular affected participant. A graph route from the participant to a competent independent reviewer is necessary for the represented appeal workflow, but not sufficient for a meaningful outcome: accessibility, mandate, fair hearing, execution, and remedy require separate assessment. Section 8 discusses these institutional properties, and the more general theory of historically generated trust belongs to the planned companion investigation.

Proposition E.3 (An appeal route is not a sufficient condition for effective review). *For a fixed claimant and a specified review procedure, directed reachability from the claimant to a review node does not imply that the claimant can obtain an effective review. Conversely, if all prescribed channels to the only competent review node are absent, review cannot be obtained through that stipulated procedure.*

Proof. For the first claim, choose a graph with a delivery path for an appeal submission but stipulate that the claimant cannot access the necessary evidence or that the review node lacks power to issue an effective decision. Reachability holds and effective review fails. For the converse, the stated procedure accepts submissions only through paths terminating at its competent review node; with no such path, no submission can arrive under the procedure. Both claims are relative to the stipulated route and competence model, not to every imaginable informal or external remedy. $\square$

Exit is similarly multidimensional. A member organization might stop new sharing, withdraw a particular service, cease participation in a shared rule-making body, or request revocation of specified derivatives. Each action can have a different effective scope. Previously published public information may remain copied, and legitimately held archival copies may be subject to retention requirements. A governance model should therefore identify which prospective exchanges an exit decision changes, which earlier obligations remain, how downstream users are notified, and who can challenge the exit conditions. Equal treatment is not guaranteed by making the same exit form available to institutions with unequal operational dependencies.

E.8 A Bounded Normative-Coordination Dossier

One further fictional case isolates the difference between agreement on data content and agreement on what to do with it. Three institutions receive the same observation record and jointly validate its metadata and calibration. They nevertheless offer three different readings of its significance and disagree about the permitted disclosure resolution. The federation should record these as independent assertions, with their respective generating methods, purposes, supporting reasons, and institutional mandates; it should not silently replace them with the ruling participant's preferred label.

A candidate *normative-coordination dossier* has the following typed fields: observation and interpretation versions; affected participants and the stated bases of standing; proposed practice rule and its bounded scope; distinct reasons for assent, rejection or reservation; legal authority if claimed; review and voting procedure; resource conditions of participation; application and performance events; dissent and reopened claims; and authorized revision or remedy events. The dossier does not decide which norm is correct. It permits researchers to reconstruct which arrangements were considered, who could affect the candidate space, and why a purportedly common rule continued or failed.

For a protocol n and documented participants P , distinguish the following predicates:

$$\text{Accept}_P(n), \quad \text{Valid}_J(n), \quad \text{Enacted}_h(n), \quad \text{Legitimate}_C(n). \quad (96)$$

Here J denotes applicable sources of legal and institutional authority, h the documented execution history, and C a declared set of normative evaluation criteria that may itself be disputed. The predicates are distinct. In particular, $\text{Accept}_P(n)$ does not entail $\text{Valid}_J(n)$, $\text{Enacted}_h(n)$, or $\text{Legitimate}_C(n)$: a unanimous but unauthorized agreement, an agreed but unimplemented rule, and an agreed rule excluding affected interests supply immediate counterexamples. Nor does disagreement automatically establish illegitimacy, because an applicable legal rule and a scientifically justified procedure can be contested. Equation (96) is a bookkeeping distinction, not a substantive legal decision procedure.

A comparative protocol should deliberately include failures: interpreted evidence that yields no workable norm; apparently successful assent under unequal dependence; valid agreements that are not enacted; and performance histories in which repeated correction makes limited cooperation possible despite continuing theoretical disagreement. These are empirical research cases for the proposed companion work on generative trust, with no assertion of universal convergence. Article 38 of the ICJ Statute and the ILC's 2026 general-principles discussion may inform an analogy about *identifying* legal sources where relevant, but cannot confer international-law status on a data-federation rule [40, 42].

E.9 Evaluation Protocol and Boundaries of Inference

The proposed governance arrangements can be compared through a structured audit, not a universal governance-efficiency score. A useful research protocol should record: the types and numbers of disputes; proportions with accessible notice and evidence; diversity of recognized claimants; provisional measures and their duration; reasons and rule versions; availability of independent review; time and resources required under a specified measurement convention; execution and downstream notification evidence; the ability to reopen decisions; and ecological or social effects not repaired by informational actions. Where participant privacy or ecological sensitivity is at issue, even aggregate publication of these metrics requires disclosure review. The relevant denominators, missing values, and excluded disputes must be stated. Rates computed only from accepted complaints would conceal potentially important differences in standing or access.

Several apparently attractive conclusions are not warranted. More logged disputes do not by themselves prove a less trustworthy federation; they may indicate greater access to remedy. Fast closure does not necessarily mean an equitable decision; it may reflect summary dismissal or limited participation. Full provenance does not establish correct science. A large number of participating institutions does not establish balanced authority. A single jurisdiction's rule cannot automatically bind every other institution participating in a data federation. Finally, a procedural remedy for mistaken disclosure need not reverse ecological harm or erase sensitive knowledge from downstream systems.

The claims in Propositions E.1, E.2, and E.3 are conditional logical counterexamples, not empirical findings about WMO, GOOS, Argo, or a particular legal regime. International programmes supply examples of coordination under distinct mandates and responsibilities, not a ready-made worldwide adjudicative order [63, 31, 57]. Future investigation would need comparative institutional fieldwork, authorized documentation of disputes, legal analysis in the relevant jurisdictions, interviews that respect confidentiality and informed consent, and carefully controlled evaluations of notification, correction, and redress workflows. The appendix's contribution is a bounded governance vocabulary: a federation can record the history of a disputed claim, make conflict explicit, and support verifiable steps toward correction, while the credibility and legitimacy of those steps remain subjects of human and institutional inquiry.

F Evaluation Metrics and Research Protocols

The proposed federation should be evaluated as a set of scientific and institutional relations whose observed performance depends on the task, sampling design, participating institutions, affected communi-

ties, and the boundary of the investigation. A successful test of network availability does not establish the validity of a scientific inference; a well-calibrated prediction does not establish permission to disseminate its inputs; and a timely governance decision does not establish an effective remedy. This appendix develops a bounded evaluation vocabulary, possible measurements, and research protocols for examining these distinctions. It extends the evaluation discussion in Section 11 and the fictional infrastructures and dispute procedures of Appendices D and E. The metrics and examples below are proposals or synthetic calculations, not measurements from WMO, Argo, GOOS, or an implemented global federation.

F.1 Evaluation Objects, Boundaries, and Estimands

An evaluation must first state what is being assessed. Let a declared evaluation protocol be the tuple

$$\Pi = (\mathcal{H}_\Pi, \mathcal{N}_\Pi, \mathcal{T}_\Pi, \mathcal{D}_\Pi, \mathcal{A}_\Pi, \mathcal{M}_\Pi), \quad (97)$$

where $\mathcal{H}_\Pi$ denotes a documented, possibly partially ordered collection of relevant observation and decision events; $\mathcal{N}_\Pi$ identifies participating and affected institutional roles; $\mathcal{T}_\Pi$ is the set of scientific and governance tasks; $\mathcal{D}_\Pi$ specifies eligible records and observation supports; $\mathcal{A}_\Pi$ records applicable permissions and responsibilities; and $\mathcal{M}_\Pi$ specifies measurements, missing-data rules, comparison conditions, and validation procedures. The protocol tuple in Equation (97) is not a commitment to an external universal clock. Where elapsed durations matter, the protocol must specify compatible timestamp conventions, synchronization tolerances, and censoring rules; otherwise relational precedence is sufficient.

For an evaluand k and task t , an estimand is a target property $\theta_{k,t}(\Pi)$, such as the proportion of eligible queries successfully completed under a given fault condition, the error of an environmental prediction on a specified reference sample, or the execution rate of a specified correction obligation. The empirical estimate $\hat{\theta}_{k,t}$ must be accompanied by its denominator, support, design assumptions, precision, exclusion rules, and provenance. A synthetic test may verify a software property within a finite test fixture, but it cannot estimate θ for an unobserved real population.

Evaluation objects should be kept separate: observation records, instrument and calibration histories, translation mappings, model inferences, delivery services, access and governance decisions, and ecological interventions. A federation-level aggregate obscures these objects unless a justified mapping explicitly links them. The FAIR principles address data findability and reuse, whereas CARE draws attention to collective benefit and authority; neither supplies a ready-made scalar score for all the scientific and normative properties considered here [61, 6].

F.2 Observation Quality, Coverage, and Scientific Reproducibility

An observation-level assessment should document reference measurement conditions, calibration traceability, units and coordinate conventions, spatial and temporal support, quality flags, missingness, and whether data are provisional or revised. Following the concern developed in Section 4, a numerical error rate is interpretable only relative to a specified reference procedure and uncertainty characterization. Argo’s real-time and delayed-mode quality-control distinction illustrates why record versions and reprocessing status matter when comparisons are made across institutions [56, 2].

For an explicitly defined finite set $\mathcal{D}_\Pi$ of required observation units, one elementary reporting-completeness measure is

$$\hat{C}_\Pi = \frac{\sum_{i \in \mathcal{D}_\Pi} w_i \mathbf{1}\{\text{an eligible, interpretable record for } i\}}{\sum_{i \in \mathcal{D}_\Pi} w_i}, \quad w_i > 0. \quad (98)$$

The weights and frame are specified by the evaluation design. Equation (98) measures completeness relative to the *declared frame*, not geographical coverage of unobserved ocean regions or independent scientific correctness. Duplicate observations need not improve it; withheld and unknown cases must not disappear from the denominator without explanation.

Proposition F.1 (Observed coverage cannot identify unobserved environmental states). *Let the available observations depend only on the restriction of an environmental field x to a proper sampled support $S \subsetneq \Omega$. In the absence of a constraint connecting the field on S to the field on $\Omega \setminus S$, no statistic of these observations alone identifies a quantity that can vary solely with $x|_{\Omega \setminus S}$.*

Proof. Construct two fields x and x' agreeing on S but differing on $\Omega \setminus S$. They generate identical observed data and every statistic measurable from those data, but a target functional depending on the difference outside S has distinct values. An identified estimate of that target from those observations would need to equal both values, which is impossible. $\square$

Proposition F.1 explains why a high reported completeness score cannot establish global representativeness. Reference comparisons must consider selection mechanisms, correlated sensor errors, model dependence, the distinction between reference data and ground truth, and changes introduced by reanalysis. Scientific reproducibility should include the ability to reconstruct a stated result from a declared combination of input versions, code or executable model identification, parameter settings, environment records, and permissible access to required resources. A result may be computationally repeatable under a particular implementation while its scientific premise remains contested.

A practical protocol should evaluate reference-consistent error when suitable independent comparison data exist; uncertainty calibration and coverage of declared prediction intervals; the reproducibility of a selected inference under explicitly described software and data versions; and the rate at which documented revisions propagate to derived results. Independent validation is needed because training or assimilating an observation into a model and then scoring the model against that same observation can overstate out-of-sample performance. Appropriate holdouts should respect instrument, geographical, institution, and relevant dependency domains. Existing assimilation practice and reanalysis documentation provide methods and cautionary examples rather than automatically portable accuracy guarantees [16, 33, 9].

F.3 Semantic Translation, Model Plurality, and Uncertainty

Translation evaluation must specify both the source representation and a *target scientific task*. For a source record r , translation T , source-task quantity Q_s , and target-task estimator Q_t , define, where an agreed comparison scale exists,

$$\delta_Q(r, T) = d_Q(Q_s(r), Q_t(T(r))). \quad (99)$$

The discrepancy δ_Q in Equation (99) is meaningful only when Q_s and Q_t have been operationally related and d_Q is defensible for that task. A semantic mapping could preserve measurement units and a mean while failing to preserve extremes, sampling supports, uncertainty assumptions, or the available inferential continuations, as Appendix C demonstrates. A zero discrepancy for one task does not establish a lossless translation.

The protocol should register known and unknown concept mappings, contextual exclusions, source versions, intended users, the validity range of each translator, and the ability to challenge mappings. Numerical comparisons should be complemented by review of category boundaries, scientific disagreements, and who has authority to declare equivalence. For the model layer, assessment should preserve distinct model versions and assumptions, and report both agreement within a given task and substantive divergence under alternative admissible models. Uncertainty sources should be separated into measurement, representational translation, parameter, structural-model, selection, and governance-related uncertainty; they cannot all be reconstructed from a single ensemble spread.

A useful counterfactual experiment creates two source datasets with identical aggregate means but different tails, then checks whether a recipient's hazard-extreme calculation changes. A second experiment uses two generative models agreeing on observed support but diverging under out-of-support perturbations. These are falsifiable, synthetic demonstrations of representation limits, not evidence that any

named international service actually commits such an error. Where target quantities lack a shared numerical scale, the protocol records structured qualitative disagreements rather than fabricating a universal distance measure. SOSA/SSN, CF conventions, and SKOS support useful descriptions and concept relations, but they do not remove these task-specific limitations [75, 7, 70].

F.4 Operational Reliability, Discovery, and Authorized Usability

An infrastructure evaluation needs distinct tests of endpoint availability, new-record discovery, successful authorized retrieval or processing, scientifically adequate interpretation, and reliable downstream correction. Let Q_{Π} be a preregistered set of task queries with both eligible and denied requests, and let $D(q)$ and $U(q)$ indicate technically deliverable and authorized/scientifically fit completion respectively. A bounded usability measure is

$$\hat{U}_{\Pi} = \frac{\sum_{q \in Q_{\Pi}} v_q \mathbf{1}\{U(q)\}}{\sum_{q \in Q_{\Pi}} v_q}, \quad v_q > 0, \quad (100)$$

under the assumption that the evaluation has verified the relevant authorization and fitness criteria. An authentication failure, deliberate and legitimate denial of a restricted request, and an unplanned outage are not interchangeable kinds of failure. For example, a restricted endpoint may correctly deny an unauthorized requester, which should count as successful enforcement of the access policy rather than unsuccessful service delivery. Equation (100) therefore applies only to the declared set of eligible tasks; policy-enforcement tests receive separate scores and explicit denominators.

The graph experiments in Appendix D show why critical discovery services, archived endpoints, caches, and review services require separate test plans. Fault injection should distinguish single-service loss, shared credentials, dependence on one cloud or power system, organizational withdrawal, and inconsistent registry versions. If two nominally redundant paths have conditionally independent availability probability p given the absence of a common-cause outage of probability q , then a simple illustrative expression is

$$P(\text{at least one path available}) = (1 - q) [1 - (1 - p)^2]. \quad (101)$$

For $p = 0.8$ and $q = 0.1$, the value is 0.864, compared with 0.96 when the common-cause mechanism is absent. These numbers are hypothetical and only characterize the stipulated probability model. Real federation evaluation requires observed dependency domains, failure correlations, reconfiguration events, and a documented distinction between discovery outages and continued access via prior knowledge.

Table 32: Proposed evaluation dimensions and distinct evidence requirements. Metrics are task-relative, not a composite score.

Dimension	Operational indicator or test	Evidence and limiting condition
Observation	Calibration and revision trace; frame-defined completeness	Reference quality, selection and missingness require investigation
Inference	Holdout performance, uncertainty coverage, reproducible derivation	Model structural errors and extrapolation remain separate
Translation	Target-task discrepancy; known unsupported mappings	Meaning and context need scientific and contributor review
Infrastructure	Discovery, delivery and usable authorized completions under fault injection	Shared failure domains and applicable permissions must be declared
Governance	Notice, standing, review, correction and verified remedy execution	Procedural record does not establish legitimacy or effective relief
Participation	Capacity to contribute, reproduce, interpret and challenge results	Resource, language and institutional asymmetry require inquiry
Ecology	Contextual energy, water, materials and environmental incidence	Baselines, counterfactuals and independent measurements required

Table 32 deliberately avoids a total ranking of global knowledge infrastructures. A lower latency or

higher replication factor can coexist with unjustified release of sensitive records, greater environmental costs, or exclusion of affected participants.

F.5 Institutional Accountability, Participation, and Remedy

Governance evaluation should begin with a register of potentially affected parties and the applicable legal and institutional mandates, not merely a set of accepted complaints. The relevant outcomes include whether an affected person or institution receives notice, has access to intelligible reasons and relevant evidence, can submit a challenge, receives a competent decision, and can verify downstream remedy execution. Even these observable components cannot replace a substantive assessment of fairness, lawfulness, scientific adequacy, or environmental responsibility. Appendix E specifies the fictional dispute procedure used as an illustrative test fixture.

A reported remedy-execution fraction requires a carefully identified population. If $\mathcal{O}_{\Pi}$ is a set of obligations triggered under a specified rule and $E(o)$ indicates independently evidenced execution within a declared evaluation window or relational closure condition, a descriptive rate is

$$\hat{R}_{\Pi} = \frac{\sum_{o \in \mathcal{O}_{\Pi}} \mathbf{1}\{E(o) = 1\}}{|\mathcal{O}_{\Pi}|}, \quad |\mathcal{O}_{\Pi}| > 0. \quad (102)$$

The rate must separately account for unknown execution, not-yet-verifiable claims, disputed obligations, and inaccessible archives. Equation (102) is undefined when no obligations were triggered; it is neither zero nor one by convention. It does not identify the harms that may persist after a technically successful correction.

Proposition F.2 (Equal observable closure rates need not imply equal access to remedy). *Two governance arrangements can display the same fraction of closed accepted disputes while differing in the proportion of affected parties able to file, obtain evidence, receive independent review, and secure executed remedies.*

Proof. Construct two fictional arrangements in which all ten accepted disputes are marked closed. In the first, ten of ten affected parties can file and all ten have evidence access and verified executed remedies. In the second, only ten of one hundred affected parties can file; those ten are summarily closed without evidence or execution. Both publish a closure fraction of $10/10 = 1$, while access and remedy differ. The constructed equality establishes non-identification from the published closure rate; it does not rank real institutions. $\square$

Proposition F.2 motivates denominators inclusive of known and potentially excluded claimants and qualitative investigation of those who cannot enter the procedure. Metrics should distinguish complaint volume, admission rates, reasons for rejection, accessibility, time and expense under a specified measurement convention, independent review, downstream notification, appeal, and documented remedy execution. Institutions should have protected ways to report suppressed complaints or unresolved grievances. Participants with fewer computational and linguistic resources may need independent technical assistance to meaningfully challenge a scientific claim. An appeal portal alone is not evidence of effective participation [52, 57].

F.6 Equitable Scientific Capacity and Recognition of Contributors

Access to a common dataset and authority to use it do not imply equal ability to reproduce or critically evaluate a published inference. A protocol should measure the required compute budget, availability of necessary software and calibration resources, access to secured processing environments, expert review opportunities, and the practical ability of less-resourced participants to propose revisions. An experiment might randomly assign access to a documented reproducibility package under appropriately consented conditions, while recording not only success but also reasons for nonparticipation, labor costs, time, and

dependence on proprietary intermediaries. Random allocation is not appropriate when it would deny rights or essential warning information; nonexperimental and participatory designs may be required.

Contributor recognition should record data stewardship, local observations, instrument maintenance, community knowledge, interpretive work, and correction efforts, without flattening these contributions into a single monetary or bibliometric unit. Collective or community authority may require access restrictions and participation rules not captured in conventional author lists. CARE and the UNESCO Open Science Recommendation offer ethical reference points; specific obligations must still be determined through applicable law, institutional practice, and participation by those affected [6, 57]. A strong protocol reports capacity differences alongside outcomes and documents who designed the evaluation criteria. It should not impose one external notion of legitimate participation as a universally binding definition.

F.7 Ecological Performance, Resource Accounting, and Counterfactuals

The proposed infrastructure consumes energy and materials and may affect land, water, and ecosystems through facilities, networks, sensors, and maintenance activities. Ecological evaluation should therefore preserve a vector of measured or estimated impacts under a clearly defined system boundary, rather than assuming a common monetary scale:

$$\mathbf{e}_{\Pi}(a) = (E_{\text{elec}}, W_{\text{withdrawal}}, W_{\text{consumption}}, M_{\text{embodied}}, L_{\text{footprint}}, H_{\text{ecological}})_a. \quad (103)$$

The entries of Equation (103) may have incompatible units, measurement supports, uncertainties, and decision relevance. A comparison across actions a and a' should specify comparable functions served, ecological baselines, location and resource context, operating horizon, upstream supply-chain boundaries, and the distribution of impacts among affected systems. The absence of a numerical value for $H_{\text{ecological}}$ must be reported as an evidence gap, not silently set to zero.

A simple impact contrast $\Delta \mathbf{e} = \mathbf{e}_{\Pi}(a) - \mathbf{e}_{\Pi}(a')$ is not a causal effect unless the design supports the counterfactual interpretation. Power-source accounting, water stress, manufacture, disposal, land conversion and habitat disturbance should be examined separately; energy-efficiency improvements can coincide with larger total consumption. The European Commission's data-centre sustainability reporting arrangements and the IEA's analyses provide external context for some measurable quantities, without supplying a complete assessment of a proposed marine or global observation federation [17, 41]. Any marine deployment, including autonomous observation systems, needs site-specific ecological review; it is not justified solely by better computation or nominal low-carbon electricity.

F.8 Comparative Study Design, Failure Conditions, and Evidence Governance

The evaluation programme should proceed through distinct, documented modes of inquiry. First, formal and synthetic tests establish logical properties of bounded schemas, graph structures, and algorithms; they require fixed versions of code and assumptions, and their conclusions are limited to the model. Second, controlled federation sandboxes test data discovery, access enforcement, computation and correction under faults and institutional rule conflicts, ideally with independent observers and safe fictional records. Third, domain-scientific evaluations compare predictions and translations against independently curated references with instrument- and geography-aware holdouts. Fourth, participatory institutional studies investigate access, trust, reasons for disagreement, rights, and remedies under consent and appropriate research ethics. Fifth, environmental assessment requires independently supported baselines, alternatives, site-specific observation, and disclosure of unresolved ecological uncertainty.

Table 33 supplies a protocol hierarchy, not a requirement that all investigations be conducted by a single institution. Each protocol needs prospective disclosure of research questions, task frames, test populations, versions, primary indicators, plausible confounders, missing-data mechanisms, governance of sensitive records, and criteria for suspending unsafe activity. Where preregistration could expose protected ecological locations or confidential institutional procedures, a versioned confidential preregistration with an appropriate independent custodian is preferable to indiscriminate publication. Independent

Table 33: Proposed research protocols, possible failures, and permitted inferences.

Protocol	Specified test and potential failure	Interpretive boundary
Formal/synthetic	Schema invariants, graph cuts, translation counterexamples, fixed-seed regressions	Tests defined model; cannot infer deployment quality
Operational sandbox	Unauthorized request rejection, shared-domain outages, version conflict and correction propagation	Requires realistic roles and dependencies for transfer
Scientific holdout	Independent calibration, out-of-domain inference, uncertainty calibration and revision reproducibility	Observational selection and reference uncertainty matter
Institutional fieldwork	Accessible filing, affected-party standing, review, execution and perceived legitimacy	Context-specific, confidential, and not captured by a score
Ecological assessment	Functional baseline, alternative comparison, life-cycle impacts and ecological response	Site, time support, causal design and uncertainty required

replication should be supported through permitted access or controlled computation rather than through unlawful copying of restricted data.

F.9 An Operational Trial Dossier and Decision-Error Benchmarks

For each test, fix a versioned trial configuration $\Pi = (Q, N, S, M, A, V)$, where Q is the set of eligible tasks and assessment questions, N the participating research nodes, S the documentary source and dependency boundary, M the candidate models and mediation procedures, A the attributed review authorities and affected-participant procedures, and V the protocol, source and software versions. These symbols are labels for evaluable components; they do not imply that legal or cultural authority is exhaustively enumerable. The test dossier must further record exclusion rules, missingness, protected-data procedures, independent reviewers and the planned conditions for rejecting, withholding or revising the evaluation.

An explicit intervention matrix can compare a basic interoperability arm a_0 , a generative-hypothesis and loss-record arm a_1 , and an arm a_2 also using competent review-support and correction tracking. Tasks must be defined so that their possible legitimate uses and known critical incompatibilities can be evaluated independently of the candidate interfaces. Where comparisons involve participating researchers, assign comparable tasks with randomized or counterbalanced order when appropriate, hold scientific resources and disclosure rights comparable across arms, and document carryover and practice effects. The study should not artificially create a difference in participants' legal entitlements, conceal consequential risks, or use private community knowledge without an independently justified basis.

For an eligible reference set Q_+ consisting of tasks assessed as legitimately usable within the study boundary, let L be the tasks for which a pilot public endpoint releases the proposed result. Conditional usable-release coverage is

$$\text{QualifiedCoverage}_{Q_+} = \frac{|L \cap Q_+|}{|Q_+|}, \quad |Q_+| > 0. \quad (104)$$

It should be presented alongside the unsafe-release rate in Equation (51), the *hold-for-review* fraction, and category-specific error counts. These rates have no defined numerical value if their reference denominator is empty and are not estimates of real-world probability without a justified sampling design. In cases of disputed reference status, report the disagreement, conditional results under alternative adjudications and reasons a numerical estimate is unavailable. Comparisons should preserve the distinction between a scientific reference decision, the right to disclose a source, and the authority to permit the downstream act.

For a source correction e and a declared, known dependency graph D_h , let $R_h(e)$ be the nonempty recorded set of descendants eligible for screening, $S_h(e) \subseteq R_h(e)$ those with documented screening, and $N_h(e)$ the contactable recipients to whom a notification is owed under the pilot protocol. Useful process

measures include

$$\text{ScreenedCoverage}_h(e) = \frac{|S_h(e)|}{|R_h(e)|}, \quad \text{NotifyCoverage}_h(e) = \frac{|N_h^{\text{notified}}(e)|}{|N_h(e)|}, \quad (105)$$

provided both denominators are nonzero and the notification obligations have been established. The first ratio measures bounded screening of recorded descendants; the second measures documented notifications to known contactable recipients. Neither measures all real-world downstream effects. The dossier must separately describe missing graph edges, unreachable recipients, disputed obligations, non-responses, verified computational changes and observed changes to real practices. A correction can alter a published claim without implying that its recipients have acted upon it.

Table 34: Measurement boundaries for the pilot evaluation; quantities should not be combined into a single infrastructure score.

Dimension	Required observations	Important non-identification
Task-relative translation	Reference-supported matches, known mismatches, unresolved cases, declines and rationales	Correct wording does not establish generative equivalence
Public release and authority	Release, hold, refusal, reviewed scope and independent source status	A signed dossier cannot confer legal competence
Correction and contestation	Known descendants, screenings, notifications, appeals and verified changes	Known dependency reach does not measure total influence
Common productive capacity	Eligible applicants, realized compute/instrument allocation, reproducible tasks and costs	Artifact access does not measure independent agency
Marine and ecological effects	Functional service units, lifecycle inventory, baseline, habitat risk and alternatives	Simulation or local energy savings do not identify field benefit

Predeclared study failure includes a public release of a known restricted record, presentation of a false mapping as tested equivalence, silent promotion of an unreviewed policy assertion into permission, loss of an in-scope correction without an alert, or exclusion of an otherwise eligible participant for an undisclosed reason. These failures are relevant to the specified implementation and require investigation; they are not proofs of a universal mathematical impossibility. An ecological safety trigger, a competent legal refusal, withdrawal of relevant consent or an independent decision that the study lacks an acceptable evidentiary baseline should suspend the affected operation. Numeric site-specific ecological thresholds and binding legal conditions cannot be chosen by a synthetic test script and require appropriate independent processes.

A small JSON dossier distributed with the editable project illustrates the *logic* of release, hold and rejection over invented statuses, including one example where a technically valid and epistemically plausible operation must be held because of an unresolved review. The associated deterministic script checks the gate's declared invariants and demonstrates how a naive acceptance rule would mishandle cases. This example has neither domain-expert ground truth nor evidence of legal validity. Its only warranted conclusion is that a transparent and locally testable interface can represent selected failure modes without granting itself institutional authority. Together with Table 21, it gives the constructive proposal a reproducible starting point while leaving its practical benefits and harms open to independent investigation.

The proposed theory would be challenged by recurrent failures to preserve declared task-specific semantics across translations, an inability to propagate corrections despite documented responsibilities, unanticipated dependence on critical central intermediaries, materially unequal ability to contest analyses despite nominal access equality, or ecological costs incompatible with the programme's stated obligations. Conversely, successful software and simulation results alone would not establish that a worldwide epistemically and institutionally legitimate federation has been achieved. This paper offers testable archi-

texture proposals and classifications; it neither implements their entirety nor substitutes computational indicators for scientific, political, and ethical judgment.

G Evidence and Source-Verification Ledger

The present investigation combines descriptions of operational scientific programmes, published technical standards, philosophical and institutional scholarship, conditional mathematical propositions, hypothetical federation designs, and small synthetic verification exercises. Those materials serve different evidential functions. An official description may document a programme’s stated roles and interfaces while leaving its actual performance open to independent evaluation; a valid theorem may follow from stipulated assumptions without determining whether those assumptions describe the world; and a passing software test may establish a particular computational property without certifying scientific accuracy, lawful access, or ethical legitimacy. This appendix organizes the boundaries among these forms of support and specifies procedures for maintaining a revisable source-to-claim ledger. Its purpose is auditability and disciplined interpretation, rather than the retrospective declaration that all claims in the manuscript have been independently verified.

G.1 Audit Object, Claim Identity, and Scope of Verification

A useful ledger treats a claim as an identifiable, versioned object rather than an undifferentiated sentence surrounded by references. Let a claim record be represented by

$$c = (\iota_c, p_c, \sigma_c, \alpha_c, \omega_c, \nu_c), \quad (106)$$

where ι_c is a stable claim identifier; p_c is the proposition, description, or evaluative statement being advanced; σ_c records its scope, relevant observation and institutional boundaries, and declared conditions; α_c is its epistemic status; ω_c identifies available warrants and the support relation assigned to them; and ν_c denotes the record version and revision links. A claim’s grammatical tense, citation count, or apparent mathematical precision does not determine α_c . A status can change when the underlying claim, cited source, or evaluation design changes. No common universal time coordinate is required for such revisions; version identifiers and a declared precedence relation suffice for the audit trail.

The ledger distinguishes at least three objects: a *source record* (a publication, standard, dataset, legal text, research protocol, or executable artifact), a *verification event* (a particular examination of an explicitly named portion of that record), and a *claim-support judgment* (the reason the examined portion bears upon the claim’s wording). Sources can have valid bibliographic identities without having been read in full; verifying a DOI or publication date establishes neither the accuracy of all content nor the support of a nearby inference. Moreover, a verified relation between one cited passage and one narrowly worded claim does not automatically transfer to larger claims assembled from several premises.

For reproducible checking, the associated verification record should identify the reviewed source version or access date, exact passage or section, verification method, reviewer role where available, claim wording checked, and unresolved objections. Restricted materials may be represented through controlled identifiers and authorized custodial records; a publicly inspectable citation is not always a permissible output. A transparent ledger can disclose the existence and limitation of such restricted evidence without disclosing the sensitive observation itself.

G.2 Evidence Categories and the Strength of a Warrant

Table 35 distinguishes the principal types of statement used throughout this paper. The categories concern *the kind of warrant required*, not an ordinal ranking of intellectual worth. A philosophical argument can be well developed without being an empirical result, while an operational service claim can be well documented without supporting a normative inference. The appropriate standard of scrutiny follows the claim’s content.

Table 35: Claim classes, admissible warrants, and inference boundaries in the present manuscript.

Claim class	Appropriate verification object	What it does not establish alone
Programme description	Dated institutional documents, operational specifications, recorded activities	Independent effectiveness, universal adoption, legitimacy
Technical standard	Versioned normative text, conformance requirements, test profiles	Semantic equivalence, scientific truth, authorized use
Scientific observation	Calibrated records, sampling design, uncertainty and independent comparison	Generality outside support, finality of interpretation
Formal result	Explicit assumptions, valid derivation, definitions and edge cases	Applicability to actual institutions or ecosystems
Synthetic demonstration	Executable code, input fixture, expected output and reproducibility log	Real-world frequency, causal impact, external validity
Institutional or legal claim	Competent authority, applicable rules, decisions, procedural evidence	Jurisdiction-independent permissions or effective remedy
Normative evaluation	Publicly stated evaluative premises, affected interests and reasoned contestation	An objective scalar verdict from technical indicators
Research hypothesis	Test design, plausible alternatives, disconfirming observations	Confirmation by conceptual plausibility or precedent

A four-part descriptive status is sufficient for a small working ledger: *documented* means that an identified source was examined for a specific bounded statement; *derived* means that a conclusion follows from stated premises or a completed computation subject to a checked argument; *proposed* identifies an architecture, model, normative framework, or test that has not been field-validated; and *unresolved* marks a claim requiring additional evidence or conceptual refinement. These are labels on particular claims, not blanket labels on publications. A single source may support a documented description and motivate a proposed inference without providing evidence that the inference has been achieved.

For clarity, verification also has levels of *depth*: identification of bibliographic metadata; examination of a publisher abstract or institutional overview; review of a relevant text or standard section; verification of a formal derivation or executable example; and independent empirical replication or institutional assessment. These steps cannot be collapsed into a universal numeric credibility score. A manuscript can have extensive citations yet only partial coverage of the source-to-claim relationships that matter most.

G.3 International Observation Programmes as Bounded Precedents

The institutional cases in Section 3 serve as precedents for particular forms of cooperation, rather than proof of the full proposed federation. WMO’s description of WIS2 identifies publication nodes, discovery functions, notification and data-transfer mechanisms, and a staged operational transition. Official WMO documents support these specific architectural and institutional descriptions, while deployment reliability, social uptake, and benefit distribution require additional independent evidence [66, 68, 64, 63]. Claims about system status must remain tied to an identified documentary version and reference period; an operational announcement must not be silently converted into a claim of universal implementation.

Argo documentation provides a different kind of precedent. Public programme materials distinguish real-time observations from delayed-mode quality-controlled records and explain the importance of adjusted data, uncertainty estimates, and flags. This substantiates the need for versioned scientific records and the possibility of retrospective correction. It does not demonstrate that every ocean observation is complete, that all derived products are error-free, or that a general-purpose global epistemic system already exists [1, 2, 56, 62]. The reported quality-control procedures and the merits of the proposed federation occupy different evidential categories.

GOOS and OceanOPS contribute evidence for international coordination of observing networks, variables, deployment and operational information under plural organizational responsibilities [58, 31, 39]. The European Digital Twin Ocean and EDITO illustrate documented initiatives relating observations,

models, and computational services, but the description of a developing platform is neither proof of fully operational worldwide coverage nor proof of scientific equivalence among participating models [20, 21, 19]. Likewise, the UN’s early-warning programme supplies a framework connecting risk knowledge, monitoring, communication, and preparedness, whereas local effectiveness depends on arrangements and evidence beyond the framework itself [59, 69].

Table 36: Bounded claims retained from the principal institutional cases. “Further evidence” identifies tasks for later research, not known deficiencies in every programme.

Case	Documented relevance to this paper	Further evidence needed for broader claims
WIS2 / WMO	Exchange infrastructure, roles, common specifications, policy categories	Coverage by context, resilience, cross-border benefit, rights and recourse
Argo	Distributed measurement, delayed quality control, record revision	Sampling representativeness, full derivative lineage, impacts on decisions
GOOS / OceanOPS	Coordinated observing variables, networks and operational metadata	Governance effectiveness, resource equity, independent performance comparison
EDITO	Developing ocean-model, data and computing integrations	Deployed interoperability, reproducibility, lifecycle ecological effects
Early warning	Multi-component science-to-action framework	Last-mile response, local participation, causal reductions in loss

The comparative reading of Table 36 supports a research strategy: identify a concrete documented mechanism, specify the extrapolation being proposed, and state the observation or institutional evaluation needed to test it. It does not support the inference that previous cooperation eliminates geopolitical competition or the need for new forms of credible commitments and governance.

G.4 Technical Standards, Semantics, and Knowledge Claims

Technical specifications provide more determinate evidence about representation and exchange than about knowledge validity. W3C provenance conventions offer a vocabulary for expressing relations among entities, activities and responsible agents; DCAT supplies catalogue descriptions; SOSA/SSN describes sensor, observation and actuation terms; OGC interfaces provide interoperable routes to certain records or sensing functions; and CF conventions address scientific data metadata [71, 74, 75, 50, 49, 7]. Their stated functions support the vocabulary developed in Appendices A and B; they do not jointly entail that any two models infer the same phenomena from a record.

The FAIR principles concern the findability, accessibility, interoperability and reuse of digital objects, including machine-actionable metadata, while the CARE principles call attention to collective benefit, authority, responsibility and ethics in Indigenous data governance [61, 6]. Their use here is complementary and deliberately limited. A FAIR-compliant description can point to legitimately restricted data; conversely, open publication of a sensitive record does not demonstrate that affected communities have granted consent or that ecological disclosure risks were assessed. UNESCO’s Recommendation on Open Science supplies a relevant international normative reference while allowing questions of differentiated access, implementation and accountable participation to remain open [57].

A representative support chain is *standard text* $\rightsquigarrow$ *conformance rule* $\rightsquigarrow$ *test evidence* $\rightsquigarrow$ *bounded interoperability claim*. An additional chain is necessary to support an epistemic inference: a valid measurement context, task-compatible semantics, a defensible inferential procedure, uncertainty assessment, and evaluation against suitable reference evidence. Section 6 and Appendix C show by formal counterexample why the first chain cannot substitute for the second.

A machine-readable schema can be validated without establishing the truth of an observation or the legitimacy of its disclosure. The example JSON Schema and fictitious coastal record in Appendix B therefore have the status of a testable *proposed data profile*; they should never be described as a completed certification against all relevant international standards. A registry’s statement of a derivative’s source is similarly distinct from independent reproduction of the scientific derivation or confirmation that redistribution is permitted.

G.5 Formal Results, Synthetic Tests, and Independence of Evidence

For mathematical propositions, the ledger should preserve definitions, assumptions, proof references, counterexamples to stronger conclusions, and applicable domains. Examples include the observation-factorization condition in Proposition A.3, the translation limits of Proposition C.1, the intermediary dependence result in Proposition 12.1, and the ecological and governance metric limits in Propositions F.1 and F.2. Their proofs examine stipulated structures; validating those structures against actual practices requires separate research.

A useful verification descriptor for a result R is

$$W_R = (A_R, P_R, X_R, E_R), \quad (107)$$

where A_R records mathematical assumptions and domains, P_R names the checked proof, X_R identifies relevant counterexamples or edge cases, and E_R contains any external empirical comparison, potentially empty. Equation (107) is an audit schema, not a scalar probability of truth. A correct proof with $E_R = \emptyset$ can remain a valid theorem, but no inference about the prevalence or practical usefulness of its conditions is warranted from that proof alone.

The manuscript's small Python scripts test specifically declared synthetic inputs for JSON Schema validity, aggregation counterexamples, graph reachability and service cuts, governance logic, sensor drift, and proposed evaluation metrics. A passing check establishes the observed output of that implementation under those inputs. It does not test all possible inputs, prove the program free of defects, independently replicate an Earth-observation system, or estimate actual cross-border scientific performance. Reproducibility reports should include input versions, tested assertions, software requirements, observed outcomes and known untested cases. These reports belong alongside the code and should be maintained when the corresponding manuscript proposition changes.

Proposition G.1 (Citation multiplicity does not establish independent corroboration). *Consider a claim c and multiple documentary records $s_1, \dots, s_n$ that all reproduce a single unverified originating assertion s_0 , with no additional independent observation or argument. The number n of citing records alone supplies no independent empirical corroboration of c .*

Proof. Construct two possible underlying states, one in which c is true and one in which c is false, while the same originating assertion s_0 and its reproductions occur in each. The visible citation set $\{s_1, \dots, s_n\}$ is identical in both states and hence cannot distinguish them. Increasing n by adding further reproductions of s_0 therefore does not establish the truth of c . Independent corroboration requires additional evidential information, not merely additional references. $\square$

Proposition G.1 concerns the dependence structure of sources; it is not an accusation about any specific institution or author. It has particular relevance to press releases, programme summaries that reproduce common project descriptions, and secondary articles citing one initial report. An evidence ledger should therefore record shared source origins where they are known. It should not report a larger bibliography as equivalent to a larger number of independent measurements.

G.6 Governance, Participation, Ecological Claims, and Restricted Evidence

A distinct ledger branch is needed for the institutional and ethical judgments in Sections 8 and 9. The existence of an access policy, appeal form, or documented complaint outcome can be checked by examining its published text and, where lawfully available, administrative records. Evidence of effective participation or remedy requires further inquiry into who could submit claims, obtain information, request review, secure enforcement, and contest adverse outcomes. Cases in Appendix E are explicit fictions, not observations of national or international dispute mechanisms. Institutional and legal conditions also vary with the responsible parties, applicable jurisdictions, and specific data uses.

Ecological claims require their own observation and causal structure. A life-cycle environmental inventory may include electricity, water, material inputs and land use, but a numerical inventory alone

does not show which ecological changes were caused by a particular infrastructure. Impact claims need declared baselines, counterfactual alternatives, spatial and relational supports, monitoring periods, uncertainty, affected species and communities, and plausible confounders. The IEA's data-centre energy outlook supplies system-level contextual estimates, not a site-specific impact assessment of the proposed federation [41]. Guidance on sensitive biodiversity observations illustrates why a source can be useful for conservation while requiring qualified publication or access protections [8]. These sources support the questions and possible measures, not a conclusion that a hypothetical marine or computing installation is ecologically compatible.

Where verification would require publishing protected observations, the ledger can distinguish a publicly inspectable claim summary from a restricted evidential dossier held by an accountable custodian. The public record should state what claim was evaluated, the general method used, the authority for restriction, what remains unverified by outside parties, and any available route for independent authorized review. It must not expose precise locations of sensitive biological resources, confidential personal records or restricted security information merely to create a technically complete public provenance trail. Such restrictions constrain independent verification and should be reported as limitations rather than invisibly omitted.

The same distinction applies to AI-assisted synthesis. A generated summary, model output or proposed semantic correspondence is a derived object whose provenance and version can be recorded. Those records do not guarantee that the generating model faithfully interpreted its sources. The audit should therefore identify substantive human scholarly responsibility, independent source checking where possible, and unresolved interpretive disagreements. The paper's AI-assistance notice specifies responsibility for the text; it cannot substitute for source-to-claim verification.

G.7 Previous Research, Intellectual Attribution, and Claim Boundaries

The present paper extends a research programme that considers money, productive relations, heterogeneous value circulation, and the interpretation of generative representations. The earlier monetary-dissolution discussion supplied motivating hypotheses about monetary dependence, infrastructural conditions and distributed clearing. The earlier *Knowledge Pressing* investigation motivated attention to generative processes rather than output-level translation alone. Both should be identified as exploratory predecessors, not independent demonstrations of the effectiveness of global knowledge infrastructures [37, 35, 34].

Existing scholarship on situated observation, collective scientific criticism, classification systems, boundary objects, and epistemic injustice provides conceptual resources for analyzing these infrastructures [32, 46, 54, 5, 23]. These works are cited for specific conceptual contributions rather than enlisted as wholesale endorsements of this architecture. Similarly, the literature on commons and polycentric governance motivates comparative institutional analysis, while the actual viability of a proposed federation depends on the conditions and evidence of its own setting [51, 52].

Scholarly attribution should be specific but proportionate. A source may inspire a research perspective, supply a conceptual distinction, document a technical procedure, or establish a result directly used by an argument. These relations have different strengths and should be described accordingly. The paper's original contributions consist principally of the synthesized epistemic architecture, task-relative conceptual distinctions, proposed formal vocabulary, conditional propositions and research-design questions. The originality claim remains provisional pending comprehensive literature review and independent scholarly assessment. Repetition of a researcher's name or the multiplication of self-citations should not be mistaken for evidential strength.

G.8 Selected Primary-Source Audit and Version Boundaries (October 2026)

A further selected audit was undertaken on 25 October 2026 to inspect the relevant portions of publicly accessible institutional source pages and standards. The purpose was to verify *bounded descriptive claims* in the institutional history, implementation proposal, and governance discussion. The companion source

package retains an itemized, machine-readable register of the reviewed claim wording, bibliography key, source address, reviewed passage or heading, documentary support, and remaining uncertainty. This inspection is narrower than an exhaustive reading of the bibliography and does not purport to replicate reported measurements or establish legal validity in a disputed case.

Some temporal distinctions are consequential. The WIS2 operational announcement specifies 1 January 2025 as the start of the operational phase [66]; the EOSC Association reports an operational federation with fourteen production-grade nodes on 8 October 2026 [14]. Those are *dated institutional claims*, not independent measurements of completeness, effectiveness, or continuous service. DataCite describes Metadata Schema 4.7 as released on 3 March 2026, while the CF catalogue distinguishes released version 1.13 (December 2025) from its subsequent working draft [12, 7]. Technical references should therefore be read as version-specific instruments rather than timeless guarantees of interoperability.

The WMO provisional 2026 Guide to Instruments and Methods of Observation requires particular care: the institution’s 5 October announcement calls the edition *provisional* and states that approval consideration would occur at the November 2026 INFCOM meeting [67]. Its guidance is pertinent to the observational scenario, but this paper does not represent that provisional version as a finally approved post-meeting edition. EDITO’s September 2026 technical update documents provider-reported development activities, including data-lake synchronization and work on publication tools; it does not independently validate all envisioned digital-twin services [19].

Standards also have distinct scopes. W3C PROV represents provenance assertions, DCAT 3 supports catalogue descriptions and version-related information, OGC API–Records supplies discovery interfaces, and ODRL supplies a policy-expression model [71, 74, 50, 73]. Neither schema conformance nor the presence of a declared policy determines scientific correctness, competent authority, legal permission, or procedural legitimacy. Data-use vocabularies and community labels provide complementary precedents for representing use conditions; the identity of an appropriate decision-maker and the permissibility of an actual operation remain to be established independently [25, 45].

Table 37: Illustrative results of the selected October 2026 primary-source audit. “Supported” denotes documentary support for the stated bounded claim, not independent real-world confirmation.

Source group	Bounded documentary finding	Residual verification task
WMO/WIS2	Operational start date and differentiated exchange-service roles supported	Independent performance and rights assessment
Argo	Near-real-time and delayed-mode revision procedures documented	Sample-level calibration and correction-lineage checks
EDITO and EOSC	Dated programme development and federation-status claims documented	Deployed capability, service equity and reliability tests
DataCite, CF, W3C, OGC	Named versions and stated technical scope documented	Conformance, scientific adequacy and authorized use
Local Contexts, GA4GH	Community labels and data-use vocabulary described	Competence, permissions and affected-party review
ICJ and ILC	Primary legal text and 2026 ILC draft conclusions identified	Jurisdiction-specific applicability; no new legal authority inferred
GOOS 2025 report	Report title and institutional reference located	Original report/full statement not directly inspected in this pass

The official UN reproduction of the ICJ Statute identifies Article 38 as concerning the legal sources applied by the Court in disputes submitted to it. The International Law Commission’s 2026 material consists of draft conclusions and commentaries rather than a treaty [40, 42]. The distinction matters when international-law analogies are used to design a federation: neither source automatically makes a project protocol legally binding. A separate source-access limitation remains for the linked 2025 GOOS status-report webpage, which could not be directly inspected during this pass; detailed programme metrics should therefore be checked from the underlying report before being cited as audited observations [38].

G.9 Ledger Maintenance, Open Verification Tasks, and Release Criteria

The source-to-claim ledger is best maintained as a living record. Its update procedure is: identify the claim and exact source passage; classify the claim and verification depth; record source provenance and dependencies; check wording against the evidence and counterexamples; document any restrictions and affected interests; record uncertainty and required further review; and assign a new version when the judgment or claim changes. A later correction should preserve the link to the earlier claim rather than simply overwrite it, because auditability requires a recognizable history of interpretation and repair.

A simple non-normative coverage statistic can count claims whose declared minimal review procedure has been completed:

$$\hat{\gamma}(\mathcal{C}; \mathcal{Q}) = \frac{\sum_{c \in \mathcal{C}} w_c \mathbf{1}\{\text{required check } \mathcal{Q}(c) \text{ documented as completed}\}}{\sum_{c \in \mathcal{C}} w_c}, \quad w_c > 0. \quad (108)$$

The finite set $\mathcal{C}$, weights w_c , and check rules $\mathcal{Q}$ must be declared in advance. Equation (108) measures recorded completion of selected verification procedures; it is not a posterior probability that the manuscript is true, and it cannot certify the trustworthiness of a global scientific institution. A high value could coexist with an important unexamined conceptual error, unreliable reference observations, or inaccessible participants. Unperformed checks should remain visible in the denominator.

Proposition G.2 (Verification coverage does not establish truth or legitimacy). *For any finite claim set and verification protocol whose checks are limited to syntactic, bibliographic, or documentary completeness, a coverage score of one does not logically entail that every claim is empirically true or institutionally legitimate.*

Proof. Choose a claim that accurately cites a programme’s announced aspiration but incorrectly concludes that the aspiration has already been independently achieved. It can pass every prescribed check of source existence, correct bibliographic metadata, and presence of a relevant institutional document. The empirical conclusion can nevertheless be false. Alternatively, a correctly recorded restriction can be challenged as illegitimate by an affected participant. In either case the declared verification procedures can all be marked complete while truth or legitimacy fails. The claimed entailment therefore does not follow. $\square$

For this release, remaining tasks include a full-text and version-specific audit of each high-consequence external source; verification of the exact language and scope of institution-specific policies; independent examination of all mathematical proofs and omitted limiting cases; systematic evaluation of the code under additional adversarial fixtures; scrutiny by relevant domain scientists of the observation and assimilation claims; comparative research on access, trust and governance with affected participants; and independent ecological evaluation of any proposed physical deployment. None of these can be replaced by compiling LaTeX or running synthetic tests.

Table 38 records practical thresholds for the subsequent revision programme. Each threshold is tied to a type of evidence, and none is intended to yield a general certification of science, governance or ethical value.

This ledger therefore preserves an essential boundary for the whole inquiry. The manuscript documents existing forms of global observation cooperation and published standards; it proposes a more general epistemic and institutional architecture; it proves selected conditional formal results and tests selected fictional examples; and it identifies questions that require future scientific, legal, social and ecological investigation. These categories may inform one another, but none substitutes for the others. The completed first draft is suitable for systematic scholarly review precisely because its claims are accompanied by the conditions and limits under which they are presently supportable.

Table 38: Evidence required before strengthening selected claims in a subsequent release.

Potential strengthened claim	Required additional inquiry	Present limitation
Global federation works in practice	Implemented testbed, independent metrics, access and failure testing	Architecture and fictional network tests only
Translation preserves scientific inference	Task-defined datasets, reference methods, uncertainty and external validation	Conditional results and synthetic counterexamples only
International governance provides effective remedy	Jurisdiction-specific examination and participatory outcome evidence	Hypothetical procedures and formal distinctions only
Ecological impact is acceptable	Site-specific baseline, life-cycle accounting, alternatives, monitoring and affected-party review	General account, no original field impact estimates
Scientific and institutional equity improves	Defined participants, comparative access and capacity evidence, contestability assessment	Proposed indicators, no intervention evaluation

References

- [1] Argo International Programme. *Argo Data*. 2026. URL: <https://argo.ucsd.edu/data/>.
- [2] Argo International Programme. *Argo Data System*. 2026. URL: <https://argo.ucsd.edu/organization/argo-data-system/>.
- [3] Argo International Programme. *Argo Missions and Objectives*. 2026. URL: <https://argo.ucsd.edu/about/mission/>.
- [4] James Bogen and James Woodward. "Saving the Phenomena". In: *The Philosophical Review* 97.3 (1988), pp. 303–352. DOI: 10.2307/2185445.
- [5] Geoffrey C. Bowker and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences*. MIT Press, 1999. DOI: 10.7551/mitpress/6352.001.0001.
- [6] Stephanie Russo Carroll, Ibrahim Garba, Oscar L. Figueroa-Rodríguez, et al. "The CARE Principles for Indigenous Data Governance". In: *Data Science Journal* 19, 43 (2020). DOI: 10.5334/dsj-2020-043.
- [7] CF Conventions Committee. *NetCDF Climate and Forecast (CF) Metadata Conventions: Released Versions and Conformance*. 2025. URL: <https://cfconventions.org/conventions.html>.
- [8] Arthur D. Chapman. *Current Best Practices for Generalizing Sensitive Species Occurrence Data*. 2020. DOI: 10.15468/doc-5jp4-5g10.
- [9] Copernicus Climate Change Service and European Centre for Medium-Range Weather Forecasts. *ERA5: Uncertainty Estimation*. 2026. URL: <https://confluence.ecmwf.int/pages/viewpage.action?pageId=111158485>.
- [10] Copernicus Marine Service. *Sources of Copernicus Marine Data*. 2026. URL: <https://help.marine.copernicus.eu/en/articles/5046704-sources-of-copernicus-marine-data>.
- [11] Ubiratan D'Ambrosio. "Ethnomathematics and Its Place in the History and Pedagogy of Mathematics". In: *For the Learning of Mathematics* 5.1 (1985), pp. 44–48. URL: <https://eric.ed.gov/?id=EJ325793>.
- [12] DataCite. *DataCite Metadata Schema, Version 4.7*. Mar. 3, 2026. URL: <https://schema.datacite.org/>.
- [13] DataCite. *Versioning: DataCite Support Documentation*. 2026. URL: <https://support.datacite.org/docs/versioning>.
- [14] EOSC Association. *Building the EOSC Federation*. 2026. URL: <https://eosc.eu/building-the-eosc-federation>.
- [15] European Centre for Medium-Range Weather Forecasts. *Fact Sheet: Reanalysis*. 2023. URL: <https://www.ecmwf.int/en/about/media-centre/focus/2023/fact-sheet-reanalysis>.
- [16] European Centre for Medium-Range Weather Forecasts. *Data Assimilation*. 2026. URL: <https://www.ecmwf.int/en/research/data-assimilation>.
- [17] European Commission. *Commission Delegated Regulation (EU) 2024/1364 on a Common Union Rating Scheme for Data Centres*. Mar. 14, 2024. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1364>.
- [18] European Commission. *European Open Science Cloud (EOSC)*. 2026. URL: https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/european-open-science-cloud-eosc_en.
- [19] European Digital Twin Ocean. *EDITO Technical Updates, September 2026*. Sept. 15, 2026. URL: <https://www.edito.eu/news/edito-technical-updates-september-2026/>.
- [20] European Digital Twin Ocean. *The European Digital Twin Ocean: About EDITO*. 2026. URL: <https://www.edito.eu/european-digital-twin-ocean-about/>.
- [21] European Digital Twin Ocean. *What Is EDITO?* Aug. 26, 2026. URL: <https://help.edito.eu/en/articles/12569275-what-is-edito>.
- [22] Geir Evensen. "The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation". In: *Ocean Dynamics* 53 (2003), pp. 343–367. DOI: 10.1007/s10236-003-0036-9.
- [23] Miranda Fricker. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, 2007. DOI: 10.1093/acprof:oso/9780198237907.001.0001.
- [24] Peter Galison. "Trading with the Enemy". In: *Trading Zones and Interactional Expertise: Creating New Kinds of Collaboration*. Ed. by Michael E. Gorman. MIT Press, 2010. DOI: 10.7551/mitpress/9780262014724.003.0003.
- [25] Global Alliance for Genomics and Health. *Data Use Ontology (DUO), Version 1.0*. Feb. 23, 2021. URL: <https://www.ga4gh.org/product/data-use-ontology-duo/>.
- [26] Global Biodiversity Information Facility. *Guide for Publishing Biological Survey and Monitoring Data to GBIF*. 2025. URL: <https://docs.gbif.org/guide-publishing-survey-data/en/>.
- [27] Global Climate Observing System. *About Essential Climate Variables*. 2026. URL: <https://gcos.wmo.int/site/global-climate-observing-system-gcos/essential-climate-variables/about-essential-climate-variables>.

- [28] Global Ocean Observing System. *Essential Ocean Variables*. 2026. URL: <https://goosocan.org/what-we-do/framework/essential-ocean-variables/>.
- [29] Global Ocean Observing System. *History of GOOS*. 2026. URL: <https://goosocan.org/who-we-are/history/>.
- [30] Global Ocean Observing System. *OceanOPS*. 2026. URL: <https://goosocan.org/who-we-are/observations-coordination-group/oceanops/>.
- [31] Global Ocean Observing System. *Who We Are*. 2026. URL: <https://goosocan.org/who-we-are/>.
- [32] Donna Haraway. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective". In: *Feminist Studies* 14.3 (1988), pp. 575–599. DOI: 10.2307/3178066.
- [33] Hans Hersbach, Bill Bell, Paul Berrisford, et al. "The ERA5 Global Reanalysis". In: *Quarterly Journal of the Royal Meteorological Society* 146.730 (2020), pp. 1999–2049. DOI: 10.1002/qj.3803.
- [34] Wanhong Huang. *Heterogeneous Value Circulation in Relational Spacetime: A Topological Formalism*. 2026. DOI: 10.17613/kw3jt-5qw23.
- [35] Wanhong Huang. "Knowledge Pressing in Evolving Relational Spacetime: Inference and Translation of Generative Representations". 2026.
- [36] Wanhong Huang. "Rethinking Relation in a Generative Relational Framework". 2026.
- [37] Wanhong Huang. "Rethinking the Spacetime-Distributed Dissolution of Money: Evolving Production Relations and Infrastructural Conditions". 2026.
- [38] Intergovernmental Oceanographic Commission of UNESCO. *The Global Ocean Observing System 2025 Status Report Reveals Both Progress and Gaps in World's Critical Ocean Observation Infrastructure*. 2025. URL: <https://www.ioc.unesco.org/en/global-ocean-observing-system-2025-status-report-reveals-both-progress-and-gaps-worlds-critical>.
- [39] Intergovernmental Oceanographic Commission of UNESCO. *Joint Centre for Oceanography and Marine Meteorology in situ Observations Programmes Support (OceanOPS)*. 2026. URL: <https://www.ioc.unesco.org/en/joint-centre-oceanographic-marine-meteorological-observing-programme>.
- [40] International Court of Justice. *Statute of the International Court of Justice, Article 38*. 1945. URL: <https://www.un.org/en/about-us/un-charter/statute-of-the-international-court-of-justice>.
- [41] International Energy Agency. *Key Questions on Energy and AI: Executive Summary*. 2026. URL: <https://www.iea.org/reports/key-questions-on-energy-and-ai/executive-summary>.
- [42] International Law Commission. *Draft Conclusions on General Principles of Law with Commentaries (2026)*. 2026. URL: https://legal.un.org/ilc/texts/1_15.shtml.
- [43] Sheila Jasanoff, ed. *States of Knowledge: The Co-Production of Science and Social Order*. Routledge, 2004. URL: <https://www.routledge.com/States-of-Knowledge-The-CoProduction-of-Science-and-Social-Order/Jasanoff/p/book/9780415403290>.
- [44] JSON Schema Organization. *JSON Schema Specification, Draft 2020-12*. 2020. URL: <https://json-schema.org/specification>.
- [45] Local Contexts. *Traditional Knowledge and Biocultural Labels*. 2026. URL: <https://localcontexts.org/labels/about-the-labels/>.
- [46] Helen E. Longino. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press, 1990. DOI: 10.2307/j.ctvx5wbzf.
- [47] Helen E. Longino. *The Fate of Knowledge*. Princeton University Press, 2002. DOI: 10.23943/9780691187013.
- [48] Sally Engle Merry. "Legal Pluralism". In: *Law & Society Review* 22.5 (1988), pp. 869–896. DOI: 10.2307/3053638.
- [49] Open Geospatial Consortium. *OGC SensorThings API Part 1: Sensing, Version 1.1*. 2021. URL: <https://docs.ogc.org/is/18-088/18-088.html>.
- [50] Open Geospatial Consortium. *OGC API — Records — Part 1: Core, Version 1.0*. 2025. URL: <https://www.ogc.org/standards/ogcapi-records/>.
- [51] Elinor Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, 1990. DOI: 10.1017/CB09780511807763.
- [52] Elinor Ostrom. "Beyond Markets and States: Polycentric Governance of Complex Economic Systems". In: *American Economic Review* 100.3 (2010), pp. 641–672. DOI: 10.1257/aer.100.3.641.
- [53] John Rawls. *Political Liberalism*. Columbia University Press, 1993.
- [54] Susan Leigh Star and James R. Griesemer. "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39". In: *Social Studies of Science* 19.3 (1989), pp. 387–420. DOI: 10.1177/030631289019003001.
- [55] UK Argo Programme. *Access Data*. 2026. URL: <https://www.ukargo.net/data/access-data>.
- [56] UK Argo Programme. *Data Quality Control*. 2026. URL: https://www.ukargo.net/data/quality_control/.
- [57] UNESCO. *Recommendation on Open Science*. Nov. 23, 2021. URL: <https://www.unesco.org/en/legal-affairs/recommendation-open-science>.
- [58] UNESCO Intergovernmental Oceanographic Commission. *The Global Ocean Observing System (GOOS)*. 2026. URL: <https://www.ioc.unesco.org/en/global-ocean-observing-system>.
- [59] United Nations Office for Disaster Risk Reduction. *Early Warnings for All (EW4All)*. URL: <https://www.undrr.org/implementing-sendai-framework/sendai-framework-action/early-warnings-for-all> (visited on 10/25/2026).
- [60] Antje Wiener. *Contestation and Constitution of Norms in Global International Relations*. Cambridge University Press, 2018. DOI: 10.1017/9781316718599.
- [61] Mark D. Wilkinson et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship". In: *Scientific Data* 3, 160018 (2016). DOI: 10.1038/sdata.2016.18.
- [62] Woods Hole Oceanographic Institution Argo Programme. *Operations: Quality Control and Calibration of Data*. 2026. URL: <https://www2.whoi.edu/site/argo/operations/>.
- [63] World Meteorological Organization. *WMO Unified Data Policy Resolution (Res. 1)*. 2021. URL: <https://wmo.int/wmo-unified-data-policy-resolution-res1>.
- [64] World Meteorological Organization. *WIS 2.0 Global Services*. 2025. URL: <https://community.wmo.int/site/knowledge-hub/programmes-and-initiatives/wmo-information-system-wis/wis-20-global-services>.
- [65] World Meteorological Organization. *WIS2 Overview*. 2025. URL: <https://community.wmo.int/site/knowledge-hub/programmes-and-initiatives/wmo-information-system-wis/wis2-overview>.

- 6275 [66] World Meteorological Organization. *WMO Information System 2.0 Will Transform Sharing of Earth System Data*. Jan. 3, 2025. URL: <https://wmo.int/media/news/wmo-information-system-20-will-transform-sharing-of-earth-system-data>.
6276
- 6277 [67] World Meteorological Organization. *Provisional 2026 Edition of the Guide to Instruments and Methods of Observation (WMO-No. 8)*. Oct. 5, 2026.
6278 URL: [https://wmo.int/media/news/provisional-2026-edition-of-guide-instruments-and-methods-of-observation-](https://wmo.int/media/news/provisional-2026-edition-of-guide-instruments-and-methods-of-observation-wmo-no-8)
6279 [wmo-no-8](https://wmo.int/media/news/provisional-2026-edition-of-guide-instruments-and-methods-of-observation-wmo-no-8).
- 6280 [68] World Meteorological Organization. *WIS2 Overview*. 2026. URL: [https://community.wmo.int/site/knowledge-hub/programmes-](https://community.wmo.int/site/knowledge-hub/programmes-and-initiatives/wmo-information-system-wis/wis2-overview)
6281 [and-initiatives/wmo-information-system-wis/wis2-overview](https://community.wmo.int/site/knowledge-hub/programmes-and-initiatives/wmo-information-system-wis/wis2-overview).
- 6282 [69] World Meteorological Organization and United Nations Office for Disaster Risk Reduction. *Early Warning Systems Reach New Heights, but*
6283 *Critical Gaps Jeopardize Global Progress*. Nov. 12, 2025. URL: [https://wmo.int/media/news/early-warning-systems-reach-new-](https://wmo.int/media/news/early-warning-systems-reach-new-heights-critical-gaps-jeopardize-global-progress)
6284 [heights-critical-gaps-jeopardize-global-progress](https://wmo.int/media/news/early-warning-systems-reach-new-heights-critical-gaps-jeopardize-global-progress).
- 6285 [70] World Wide Web Consortium. *SKOS Simple Knowledge Organization System Reference*. Aug. 18, 2009. URL: [https://www.w3.org/TR/skos-](https://www.w3.org/TR/skos-reference/)
6286 [reference/](https://www.w3.org/TR/skos-reference/).
- 6287 [71] World Wide Web Consortium. *PROV-O: The PROV Ontology*. Apr. 30, 2013. URL: <https://www.w3.org/TR/prov-o/>.
- 6288 [72] World Wide Web Consortium. *Shapes Constraint Language (SHACL)*. July 20, 2017. URL: <https://www.w3.org/TR/shacl/>.
- 6289 [73] World Wide Web Consortium. *ODRL Information Model 2.2*. Feb. 15, 2018. URL: <https://www.w3.org/TR/odrl-model/>.
- 6290 [74] World Wide Web Consortium. *Data Catalog Vocabulary (DCAT) – Version 3*. Aug. 22, 2024. URL: <https://www.w3.org/TR/vocab-dcat-3/>.
- 6291 [75] World Wide Web Consortium and Open Geospatial Consortium. *Semantic Sensor Network Ontology*. Oct. 19, 2017. URL: [https://www.w3.](https://www.w3.org/TR/vocab-ssn-2017/)
6292 [org/TR/vocab-ssn-2017/](https://www.w3.org/TR/vocab-ssn-2017/).