DeSciCoLab
Decentralized Science Collaboration Laboratory

D17: Literature Review and Market Analysis Report

EXCELLENCE/0524

Project Title	Decentralized Science Collaboration Laboratory
Acronym	DeSciCoLab
Deliverable ID	**D17**
Deliverable Title	**Literature Review and Market Analysis Report**
Lead Beneficiary	University of Nicosia Research Foundation (UNRF)
Participating Beneficiary	INNOV-ACTS Ltd
Due Date	Month 3
Version	v2 (final version)
Confidentiality	Public
Author	UNRF
Quality Review	INNOV-ACTS Ltd

Co-funded by the European Union
Republic of Cyprus
RESEARCH & INNOVATION FOUNDATION
UNRF UNIVERSITY of NICOSIA RESEARCH FOUNDATION
Innov-acts

List of Abbreviations and Acronyms

DeSci: Decentralized Science

DeSciCoLab: Decentralized Science Collaboration Laboratory

Table of Contents

1 Executive Summary....6
2 Introduction and Objectives....7
2.1 Conceptual Foundations & Definitions....7
3 Methodology....10
3.1 Research & analytic questions....10
3.2 Literature review method....10
3.3 Market analysis method....11
3.4 Limitations....12
4 Literature Review....13
4.1 State of the art in decentralized research collaboration....13
4.2 Identity, attribution, and reputation systems....14
4.3 Data management and provenance....15
4.4 Decentralized storage and compute....16
4.5 Coordination, governance, and funding....17
4.6 UX patterns for researcher workflows....18
4.7 Security, privacy, and compliance considerations....19
5 Market Analysis....21
5.1 Market definition and segmentation....21
5.2 Comparative framework....22
5.3 Landscape overview....24
5.4 Adoption signals and traction metrics....25
5.5 Risks, constraints, and externalities....28
6 Comparative Analysis & Findings....30
6.1 SWOT analysis by segment....30
6.2 Unmet needs....31
7 Key insights and implications....33
7.1 Identity-first, DID optional....33
7.2 Provenance by default....33
7.3 Content-addressed storage with durable commitments....33
7.4 Compute to data for governed datasets....34
7.5 Portable assessment objects....34
7.6 Incentives tied to verifiable outcomes....34
7.7 User experience and integrations....34
7.8 Governance transparency....35
7.9 Open APIs and schemas....35
7.10 Segment-specific adoption signals....35

7.11 Compliance and privacy by design..... 35
7.12 Maintain current rubric weights for the first design cycle.....36
8 Conclusion..... 37
9 References & Bibliography..... 39
10 Appendix A. Literature review search strings.....42
11 Appendix B. Data extraction codebook and samples..... 44
12 Appendix C. Market scoring rubric..... 47
13 Appendix D. Detailed comparison tables and provider profiles.....49
14 Appendix E. Glossary and abbreviations.....57

List of Tables

Table 1: Market Layer Focus Areas and Mapped Topics 21
Table 2: Funding communities and BioDAOs 26
Table 3: Data and knowledge infrastructure 26
Table 4: Attribution, review, and integrators 27
Table 5: Storage and compute substrate 27
Table 6: Literature–Market Segmentation Matrix 30
Table 7: Market scoring rubric 47
Table 8: D1. Funding communities and BioDAOs 49
Table 9: D1. Adoption signals for D1 49
Table 10: VitaDAO 50
Table 11: HairDAO 50
Table 12: ValleyDAO 50
Table 13: AthenaDAO 50
Table 14: PsyDAO 50
Table 15: Bio Protocol (BIO) 51
Table 16: LabDAO 51
Table 17: D2. Data and knowledge infrastructure 51
Table 18: Adoption signals for D2 51
Table 19: OriginTrail DKG 52
Table 20: Galeon 52
Table 21: Open Research Knowledge Graph (ORKG) 52
Table 22: D3. Attribution and integrators 52
Table 23: Adoption signals for D3 52
Table 24: ResearchHub and ResearchCoin (RSC) 53
Table 25: OpenAlex 53

Table 26: Crossref 53
Table 27: DataCite 53
Table 28: OpenReview 54
Table 29: PREreview 54
Table 30: D4. Storage and compute substrate 54
Table 31: D4. Adoption signals for D4 54
Table 32: IPFS 55
Table 33: Filecoin 55
Table 34: Arweave 55
Table 35: Ocean Protocol (compute to data) 55
Table 36: Bacalhau (compute over data) 55
Table 37: Akash Network (decentralized compute) 55

1 Executive Summary

This report surveys the state of decentralized science (DeSci) and maps the market of projects and protocols that can support verifiable, portable research collaboration. The report serves three purposes. First, it synthesizes literature on identity and attribution, provenance and packaging, storage and compute, governance and incentives, user experience, and security and compliance. Second, it segments the market into functional layers that align with the research lifecycle. Third, it distills evidence into requirements that will guide stakeholder validation in D18, interaction design in Work Package 4 (Wireframes Development), and the technical specification and roadmap in Work Package 5.

The DeSci market is defined around four segments that interoperate through open standards. Funding communities and BioDAOs coordinate research agendas and grants in specific domains, for example longevity, women's health, synthetic biology, and psychedelic science. Data and knowledge infrastructure projects provide provenance aware graphs, health data platforms, and consent informed data flows suitable for compute to data patterns. Attribution and integrators include scholarly graphs and incentive platforms that expose programmatic interfaces and persistent identifier alignment. Storage and compute substrates offer content addressing and durable persistence, together with reproducible or attestable execution paths.

The report establishes a comparative framework uses six criteria with transparent weights. Governance and decentralization assess openness, accountability, and resilience. Identity and interoperability focus on person and object identifiers, schema alignment, and verifiable claims. Data and compute verifiability examine FAIR alignment, provenance, packaging, and execution records. Incentives and sustainability evaluate program clarity, funding stability, and anti gaming features. UX and adoption assess onboarding, documentation, integrations, and observed use. Compliance and intellectual property review licensing, privacy, and data protection controls. Initial scoring classifies projects as Leading, Advancing, Emerging, or Nascent, and links each score to public evidence.

For WP4 and WP5, the guiding principle is that verifiability should be a by product of ordinary work. A DeSci collaboration platform should offer ORCID-first sign-in with optional DID binding, auto emit provenance at deposit and execution, model reviews and replications as portable objects, and integrate with repositories, scholarly graphs, and compute orchestrators through stable APIs. We report strengths, weaknesses, opportunities, and threats by segment, and align them with adoption signals and design choices. The appendices provide the protocol, rubric, comparison tables, and glossary needed to reproduce and update this assessment.

This deliverable identifies more than ten key insights from the literature and market analysis that define the foundations for decentralized research collaboration. These insights underpin the design and technical direction of the platform developed in WP4 and WP5.

2 Introduction and Objectives

Decentralized Science (DeSci) seeks to remedy structural inefficiencies in the research lifecycle, including fragmented identity and attribution, weak and non portable provenance, limited reproducibility, and high frictions in access to data and compute. The promise is that blockchain-based networks, verifiable data structures, and programmable governance can reduce coordination costs while improving trust in results across preprint, peer review, and reuse.

In this context, this report has three objectives. First, to survey and synthesize the academic literature that bears on decentralized collaboration in science, including work on persistent identifiers, verifiable credentials, provenance models, storage and compute architectures, and governance and incentive mechanisms. Second, to map the market landscape of relevant providers and protocols, and to compare their capabilities against a transparent rubric. Third, to distill evidence that will inform stakeholder validation in D18, the interaction design in Work Package 4, and the technical specification and roadmap in Work Package 5.

The scope of this report is defined by the needs of researchers and research supporting staff who must collaborate across institutional boundaries while maintaining credit, compliance, and reproducibility. In scope are identity and attribution, data and workflow stewardship, decentralized storage and verifiable compute, mechanisms for assessment and funding, and user experience patterns that minimize overhead for busy researchers. Out of scope are low level smart contract details that do not materially influence user journeys, and speculative applications that lack methods or evaluation. We also recognize that privacy, ethics, and legal frameworks set hard constraints that a viable system must respect, for example the General Data Protection Regulation and licensing requirements for research artifacts (EUR Lex, 2016).

Methodologically, the report combines a scoping review with systematic elements. Search strategies cover Scopus, Web of Science, IEEE Xplore, the ACM Digital Library, and curated Google Scholar, together with arXiv, SSRN, OSF, and standards bodies. Evidence from the literature is integrated with a market analysis that segments the value chain, profiles representative providers, and applies a rubric that covers governance, identity and interoperability, data and compute verifiability, incentives and sustainability, user experience and adoption, and compliance and intellectual property.

The contribution of this report is twofold. It offers a synthesized view of state of the art concepts and technologies that are directly relevant to decentralized research collaboration. It also provides a decision support frame that enables reasoned trade offs in later work packages.

2.1 Conceptual Foundations & Definitions

Decentralized Science (DeSci) refers to the application of decentralized and verifiable infrastructure to scholarly workflows in order to improve attribution, provenance,

reproducibility, discovery, and funding. Open Science is the broader movement that promotes findable, accessible, interoperable, and reusable research objects, supported by community norms and infrastructure (Wilkinson et al., 2016). Reproducibility denotes the ability to obtain the same results using the same data and methods, while related notions include replicability, robustness, and transparency across the research lifecycle (Goodman, Fanelli, & Ioannidis, 2016).

Provenance is a machine readable account of entities, activities, and agents that produced a result, including inputs, transformations, and outputs. The W3C PROV family defines a conceptual model and ontology for representing such lineage in a way that is queryable and portable across systems (W3C, 2013). Research objects bundle data, code, configuration, environment, and metadata into a validated package.

Attribution and identity rely on **persistent identifiers** for people and works. ORCID provides a globally unique identifier for researchers, and DOIs provide persistent identifiers for publications and datasets. **Decentralized Identifiers** and **Verifiable Credentials** extend this stack with cryptographically verifiable identifiers and attestations that can express claims about authorship, peer review, curation, or workflow execution, with support for selective disclosure and revocation status (ORCID, n.d.; W3C, 2022; W3C, 2025).

Public goods funding denotes mechanisms that finance shared infrastructure and knowledge where exclusion is impractical or undesirable. **Quadratic funding** aggregates small signals of value through a matching function, while **retroactive public goods funding** allocates resources after impact is demonstrated.

Decentralized storage and **verifiable compute** supply integrity and assurance. Content addressing provides fixity, and networked storage markets add economic durability, with Filecoin as a prominent example. IPFS provides content addressed transport that pairs naturally with persistent identifiers, while Arweave proposes an endowment model for long term persistence. Compute can be replayable through containerized pipelines with pinned dependencies, or attestable through trusted execution environments with hardware supported measurements. In sensitive contexts the compute to data pattern keeps raw data in controlled locations and moves computation to the data, as seen in standards work in genomics and in privacy preserving collaborations such as DataSHIELD (Benet, 2014; Protocol Labs, 2017; Williams et al., 2018 or 2023; Costan & Devadas, 2016; GA4GH Cloud Work Stream, n.d.; DataSHIELD Collaboration, n.d.).

Coordination and governance mechanisms organize attention and resolve disputes. Reviews and replications can be modeled as portable objects that are signed and citable. Governance can combine off chain deliberation with on chain finality, with delegated roles, explicit remits, and transparent records. Governance defines who decides, how they decide, and how decisions are recorded and appealed. In decentralized research contexts, credible neutrality, transparency, and resilience are paramount. Token weighted voting is one option, although it often concentrates power. Contribution or reputation weighted voting can better align authority with epistemic competence, especially for decisions that affect scientific validity, such as retractions. Practical configurations combine off chain deliberation and evidence gathering with on chain finality for

records and budgets, maintain delegated roles with clear remits, and include emergency powers with time bound mandates that are reported through public post mortems. Participation thresholds and quorum rules should be tuned to avoid apathy and capture. Treasury policies must be explicit, with diversification, spending caps, and runway targets.

Incentive design ties rewards to verifiable outcomes, for example successful replications or dataset reuse, using quadratic and retroactive funding where appropriate. Incentives influence where effort is invested. Public goods funding mechanisms can support activities that matter for reliability and reuse, including curation, replication, and tooling. Quadratic funding amplifies small signals by matching community contributions, while retroactive public goods funding rewards demonstrated impact (Buterin, Hitzig, & Weyl, 2018; Goodman, Fanelli, & Ioannidis, 2016).

User experience primitives complete the picture. Authentication begins with familiar credentials, guidance is contextual, and integrations allow researchers to keep their tools while gaining verifiability.

3 Methodology

3.1 Research & analytic questions

This section summarizes the questions that guide this report. We focus on two literature questions that cover architectures, governance, and user experience patterns, and that examine evidence on effectiveness and adoption. We also focus on two market questions that cover value chain composition and comparative capability. Appendix A contains the registered search strings and source settings. Appendix B contains the data extraction codebook.

Literature questions

RQ1 (concepts and designs). What architectures, governance arrangements, incentive mechanisms, provenance models, identity and attribution methods, and user experience patterns are proposed or evaluated for decentralized research collaboration, and how do these elements relate to the quality and adoption of scholarly workflows.

RQ2 (evidence and outcomes). What empirical or quasi-empirical evidence exists regarding effectiveness, reliability, reproducibility, adoption, and limitations of the approaches identified in RQ1, and under what contextual conditions do these approaches perform well or poorly.

Market Analysis questions

AQ1 (landscape and segmentation). Which providers, protocols, and tools populate the DeSci value chain, including collaboration and review, identity and attribution, data and provenance, storage, compute, governance and funding, and integrators and tooling, and what roles do they claim within that value chain.

AQ2 (comparative capability). How do providers compare when assessed with the rubric (see Appendix C), with attention to governance and decentralization, identity and interoperability, data and compute verifiability, incentives and sustainability, user experience and adoption, and compliance and intellectual property.

RQ1 and RQ2 are addressed through a systematic literature review (see section 4). AQ1 and AQ2 are addressed through a structured market scan and comparative scoring (see section 5). All evidence is logged and extracted using the fields and rules in Appendix B, so that findings can be traced to sources and reused further in the project (D18, WP4, and WP5).

3.2 Literature review method

We use a scoping, systematic hybrid method for planning, identification, screening, inclusion, and reporting (Tricco et al., 2018). The protocol defines concepts of interest, sources, search strings, eligibility criteria, screening rules, extraction variables, and quality considerations.

Operational details are summarized here, and the registered strings and filters are recorded in Appendix A, while the extraction codebook appears in Appendix B.

The academic sources we used are Scopus, Web of Science, IEEE Xplore, the ACM Digital Library, and a curated slice of Google Scholar that focuses on the first two hundred results. Grey literature includes arXiv, SSRN, OSF, standards bodies, registries, and provider documentation. The temporal window is 2016 to the present.

We included all sources that address decentralized collaboration in research, identity and attribution, provenance and research object packaging, decentralized storage and compute in research contexts, governance and incentive mechanisms for review or replication, and user experience or evaluation studies pertinent to these themes. We excluded non-research opinion without methods, purely speculative pieces, and superseded versions where a stable successor exists.

Included items were extracted into the fields in Appendix B, which cover metadata, constructs and mechanisms, methods and evidence, outcomes and findings, identity and interoperability signals, provenance and packaging, storage and compute assurances, governance and incentives, user experience and adoption, compliance, and intellectual property.

We used narrative synthesis, supported by comparison tables and evidence linked summaries. Where feasible we cluster concepts and technologies by function and map them to requirements themes. All extractions are traceable to sources and to the findings in section 6 and the design implications in section 7.

3.3 Market analysis method

The market analysis identified providers, protocols, and tools across the DeSci value chain, and evaluated them with a transparent rubric. Segmentation followed seven functional areas, which are collaboration and review, identity and attribution, data and provenance, storage, compute, governance and funding, and integrators and tooling. The provider profiles are recorded in Appendix D.

For each provider we collected documentary evidence that includes specifications, architecture and governance descriptions, standards and schema references, application programming interface documentation, audits where available, integration catalogs, and adoption signals that include case studies and usage metrics when public.

We then performed comparative scoring using the rubric in Appendix C. The six criteria were governance and decentralization, identity and interoperability, data and compute verifiability, incentives and sustainability, user experience and adoption, and compliance and intellectual property. Each criterion is rated from one to five using the anchors in Appendix C. The composite score S equals the sum of weights multiplied by ratings across the six criteria, with weights that sum to one.

3.4 Limitations

As all methods, ours is subject to several limitations. Selection bias is possible because the most visible initiatives are often those with active communication or marketing. Language and time window introduce coverage bias because searches are limited to English and primarily to the period from 2016 onward. Grey literature varies in quality, and some documents are not peer reviewed. Vendor documentation can overstate capabilities or omit unfavorable details. The landscape is evolving quickly, which can lead to rapid obsolescence of features or claims. Finally, cross domain generalization is limited because norms and constraints differ between disciplines, for example between genomics and computer science.

We mitigated these issues through protocol registration in Appendix B, transparent exclusion reasons, and conservative ratings when evidence was ambiguous. Readers should interpret findings as a snapshot that is suitable for design and specification work in this project, with the expectation of periodic updates.

4 Literature Review

4.1 State of the art in decentralized research collaboration

Decentralized Science (DeSci) has emerged as a response to long-standing frictions in scholarly communication: fragmented scholarly identity, weak attribution of granular contributions, limited reproducibility, and constrained access to compute, storage, and funding. The core proposition is that verifiable data structures and programmable governance can realign incentives and reduce transaction costs across the research lifecycle: from hypothesis formation and preregistration through data collection, analysis, peer review, and downstream reuse. (Wilkinson et al., 2016) (W3C, 2013) (W3C, 2022; W3C, 2025)

Existing efforts can be grouped into four overlapping streams:

a) First, infrastructure primitives anchor trust and portability. Persistent identifiers for people and objects (e.g., ORCID for researchers, DOIs for works) are increasingly complemented by decentralized identifiers (DIDs) and verifiable credentials (VCs), enabling portable, attestable claims about authorship, review activity, dataset stewardship, and method execution. These claims can bind to research objects and workflows with cryptographic guarantees, supporting machine-actionable provenance and audit trails. (ORCID, n.d.) (W3C, 2022; W3C, 2025)

b) Second, data and workflow stewardship has moved from narrative description to structured, FAIR-aligned publication of research objects, including standardized metadata, persistent storage, and versioned computational artifacts. Provenance models (e.g., W3C PROV) and packaging profiles (e.g., RO-Crate) provide a lingua franca for tracing how results are produced and reused. The shift from PDFs to linked artifacts is partial and uneven, but the direction is clear: reproducibility requires executable context, not merely textual exposition. (Wilkinson et al., 2016) (Soiland-Reyes et al., 2022) (W3C, 2013)

c) Third, coordination and review are being re-imagined. Reputation-bearing review systems, public review markets, and bounty mechanisms have been piloted to improve signal quality and timeliness. A notable design pattern is the separation of assessment (the evaluative act) from venue (journal/platform), so that attestations are portable across venues and composable with other credentials (e.g., replication badges, dataset curations). (Review Commons, n.d.)

d) Fourth, funding and incentives have diversified beyond traditional grant cycles. Quadratic funding, retroactive public-goods funding, and impact-claim mechanisms attempt to reward valuable but historically under-compensated activities (curation, replication, tooling). Early evidence suggests that incentive design needs tight coupling to verifiable signals (e.g., successful replications or dataset reuse) to avoid gaming and to ensure durability beyond market cycles. (Buterin, Hitzig, & Weyl, 2018) (Optimism Collective, n.d.) (Hypercerts Foundation, 2023)

However, despite progress, gaps persist. Identity remains split between scholarly PIDs and Web3 identities; provenance is unevenly implemented outside data-intensive subfields; and many DeSci platforms struggle to escape the bootstrapping problem of two-sided markets (insufficient reviewers/users until value is apparent). Moreover, governance maturity varies: some systems remain effectively centralized behind multisigs or foundations, raising questions about credible neutrality for contentious decisions (e.g., retractions, disputes over authorship). Finally, UX frictions (key management, onboarding, and integration with incumbent scholarly tooling)remain a first-order barrier to adoption.

4.2 Identity, attribution, and reputation systems

Identity and attribution are the substrate on which scientific credit, accountability, and incentives run. Today's scholarly stack centers on ORCID (person-level identity), DOIs (object-level identity), and emerging contribution taxonomies such as CRediT. DeSci extends this by introducing DIDs (identifier method independent of any single registrar) and VCs (tamper-evident, selectively disclosable claims) to express fine-grained attestations: "X performed statistical validation on dataset Y," "Z conducted peer review for preprint W," or "Q executed workflow V in environment E with checksum C." These attestations become portable across platforms and time, and, if signed by recognized issuers, machine-verifiable. (ORCID, n.d.) (W3C, 2022; W3C, 2025) (Brand, Allen, Altman, Hlava, & Scott, 2015)

A mature attribution layer must satisfy three properties. Continuity requires that a researcher's identity spans contexts: institutional SSO, ORCID, and DID should be linkable without exposing more personal data than necessary. Composability requires that claims can be aggregated (e.g., multiple VCs forming a contribution record) and mapped to existing scholarly graph edges (author, work, reviewer, manuscript, curator, dataset). Interoperability requires alignment on schemas and APIs so that identity and attribution travel with the research object, not with the hosting platform. In practice, this points to a hybrid model: ORCID (as the canonical scholarly PID) remains the user-facing anchor, while DIDs/VCs provide cryptographic transport and selective disclosure for high-stakes attestations (ethics approvals, data-use permissions, reviewer attestations).

Attribution also intersects with **reputation**. Portable review and replication credentials can underwrite non-financial rewards (visibility, invitations) and financial ones (bounties, retroactive grants). However, portability raises adversarial concerns: sybil resistance, credential inflation, and reputation laundering across venues. Three mitigations are standard: (i) verifiable issuer registries and allowlists (who is permitted to issue which claims); (ii) revocation/expiration mechanisms; and (iii) context-sensitive scoring that discounts low-information claims.

The literature also cautions against over-indexing on cryptography at the expense of ergonomics. Researchers should be able to authenticate with familiar credentials (e.g., ORCID OAuth) and obtain a DID behind the scenes; credential wallets can be abstracted behind institutional or

cloud agents, with export options for advanced users. Interoperability with incumbent systems, for example, editorial platforms, repository deposit workflows, and research information systems (CRIS/RIMS), could be a major precondition for adoption.

4.3 Data management and provenance

Reproducibility depends on more than access to data; it requires verifiable lineage from research questions through to results. Contemporary best practice is articulated in the FAIR principles (findability, accessibility, interoperability, reusability), enriched by machine-actionable metadata and stable identifiers for datasets, code, workflows, and runtime environments. The provenance problem is to encode who did what, when, with which inputs, producing which outputs, in a form that is queryable and robust to platform churn. The W3C PROV family (PROV-DM/PROV-O) provides a well-studied graph model for entities, activities, and agents; packaging profiles such as RO-Crate bind this graph to files, containers, and metadata manifests so that research objects can be validated, migrated, and recomposed. (Wilkinson et al., 2016) (W3C, 2013) (Soiland-Reyes et al., 2022)

DeSci's contribution is to add cryptographic integrity and, where appropriate, economic durability. Content addressing (hashing) ensures that identical artifacts resolve to identical identifiers; storage networks supply durability and availability; append-only ledgers and notarization services timestamp key lifecycle events (e.g., preregistration, dataset release, analysis run). Verifiable compute contributes two complementary assurances: replayable pipelines (containerized workflows with pinned dependencies) and attestable execution (e.g., TEE attestations or, prospectively, zero-knowledge proofs for specific classes of computation). In many domains, replayability is sufficient; fully attestable compute is reserved for cases where confidentiality or high-assurance verification is mandatory. (Costan & Devadas, 2016)

The major adoption risk is **overhead**. If provenance capture requires manual curation, it will decay. A practical design automatically emits provenance as a by-product of ordinary work: deposit to a repository triggers PID minting and RO-Crate assembly; running an analysis in an instrumented environment emits a signed execution record; publishing a preprint attaches resolvable references to the underlying artifacts. The second risk is **fragmentation**: multiple partial standards and repository practices can strand metadata. Hence the importance of mapping to a small set of canonical profiles and publishing crosswalks for conversion.

Finally, **rights and ethics** are integral to data management. Data-use governance (licenses, consent constraints, data-processing agreements) should be captured as first-class metadata and, when feasible, enforced by access-control policies that travel with the object (e.g., encrypted artifacts with policy-based decryption). This aligns compliance with usability: the same descriptors that enable discovery also gate lawful reuse.

4.4 Decentralized storage and compute

At the storage layer, decentralized systems promise integrity, durability, and resistance to single-point failure by shifting from location- to content-addressed retrieval. In practice, three patterns dominate. First, gossip-based transport with pinning (e.g., IPFS-style) offers verifiable addressing and efficient deduplication but depends on explicit pinning or third-party pinning services for persistence; without clear pinning policies, availability decays over time. Second, market-based durability (e.g., proof-of-replication/proof-of-spacetime models) couples storage to economic commitments; durability can be monitored on-chain, and repair mechanisms are incentivized, but retrieval latency, retrieval pricing, and market cyclicality must be managed. Third, endowment-funded "permaweb" models front-load costs via a one-time payment for persistence; this simplifies budgeting but shifts risk to the endowment's long-run sustainability and to gateway concentration. (Benet, 2014) (Protocol Labs, 2017)

Across these patterns, research-grade stewardship introduces additional requirements: (i) **bind artifacts to persistent identifiers** (PIDs/DOIs) and fixity digests so that content addressing complements scholarly identifiers; (ii) **package metadata and artifacts** using community profiles (e.g., RO-style crates) so that datasets, code, and run environments are resolvable as a coherent object; (iii) support **versioning** with explicit lineage (parents/derivatives) to enable precise citation and rollback; and (iv) publish **discoverability metadata** (discipline, method, license, data-use constraints) sufficient for indexing and compliance checks. Integrity and durability are only meaningful if discoverability and legal reusability are first-class.

Compute introduces a different set of trade-offs between reproducibility, assurance, and confidentiality. A pragmatic baseline is replayable pipelines: containerized workflows with pinned dependencies and seeded randomness, scheduled on decentralized or commodity clouds, and emitting signed run records that capture inputs, parameters, and outputs. Replayability provides strong practical reproducibility at comparatively low cost, particularly when data can be cached near compute. Above this baseline sit attestable execution models. Trusted Execution Environments (TEEs) can offer hardware-backed attestations that a job ran in an enclave with a particular measurement, improving confidentiality but introducing supply-chain and vendor trust assumptions. Cryptographic verification, either in the form of interactive verification or succinct proofs of computation, increases assurance and composability but is currently tractable only for specific classes of workloads or at significant proving cost. A third class, optimistic verification, relies on redundancy or fraud-proof style challenges with economic penalties; it reduces cryptographic overhead but requires careful adversary modeling. (Costan & Devadas, 2016)

Two systemic frictions recur in applied research settings. First, data governance: sensitive data often carry jurisdictional or consent constraints that require compute-to-data topologies, policy-enforced decryption, and auditable access. Second, nondeterminism: floating-point behavior, multi-threading, and hardware variance can produce divergent outputs even with pinned environments; mitigation includes deterministic libraries, container image provenance,

and explicit tolerance envelopes in evaluation. Finally, cost and performance management are material: decentralized markets can be price-efficient but volatile; hybrid architectures (decentralized storage, federated compute, institutional cache) often deliver better total cost of ownership for research pipelines.

4.5 Coordination, governance, and funding

Scientific collaboration is, at root, a coordination problem: aligning attention, effort, and resources under uncertainty. DeSci reframes this problem with programmable institutions that can separate **assessment** from **venue**, reward non-publication contributions (curation, replication, tooling), and increase transparency in prioritization and allocation. Three design domains warrant careful treatment.

Coordination and review. Traditional peer review centralizes editorial discretion and often obscures reviewer labor. Emerging systems experiment with portable assessments (reviews as signed, citable objects), open review (identity and dialogue made public), and market-like mechanisms (bounties for targeted evaluations, contests for replication). The central risks are adverse selection (low-quality participation), collusion, and grade inflation under poorly designed incentives. Mitigations include calibrated scoring rubrics, reviewer reputation that travels via verifiable credentials, randomized assignment within expertise clusters, and ex post audits (e.g., replication spot checks). Importantly, portability does not imply anarchy: editorial functions (scope, standards, conflict resolution) remain, but their decisions become auditable artifacts rather than opaque gatekeeping. (Review Commons, n.d.)

Governance. Token-based voting is neither necessary nor sufficient for credible governance. In practice, viable systems adopt **hybrid models**: off-chain deliberation with on-chain finality; delegated authority with explicit remits; emergency powers under strict sunset; constitutional constraints on treasury use and parameter changes; and mandatory **disclosure of conflicts**. Voting rights may be tied to **reputation or contribution** rather than to capital alone, and quorum/participation thresholds should be tuned to avoid plutocracy and apathy. Where decisions affect scientific validity (e.g., retractions), **credibly neutral** processes with appeal paths are essential to maintain trust beyond the immediate community.

Funding and incentives. Public-goods underprovision is acute in research: replication, curation, and infrastructure are essential but historically under-rewarded. Mechanisms such as quadratic funding (ex ante) and retroactive public-goods funding (ex post) attempt to amplify small, distributed signals of value and to pay for realized impact rather than promises. Impact-claim schemas (e.g., semi-fungible claims bound to projects) allow multiple funders to recognize the same contribution without double counting, but they require well-specified impact oracles (how impact is measured), anti-sybil measures, and governance over disputes. Treasury management should be transparent and conservative (for example, through diversified reserves, clear runways, and contingency planning) so that incentive programs persist across market cycles.

Finally, the literature emphasizes coupling incentives to verifiable signals (successful replications, dataset reuse, code adoption) to reduce gaming and align rewards with epistemic quality. (Buterin, Hitzig, & Weyl, 2018) (Optimism Collective, n.d.) (Hypercerts Foundation, 2023)

For all three domains, the **interface with identity and provenance** is decisive: assignments, attestations, and grants must bind to people and artifacts in ways that survive platform churn. Equally, **UX matters**: cryptographic ceremony should be hidden behind familiar actions (submit, review, replicate), and the platform should integrate with researchers' existing tools rather than demand workflow replacement.

4.6 UX patterns for researcher workflows

User experience is decisive in determining whether decentralized architectures are adopted or bypassed. The literature and practice suggest that "cryptography-forward" designs underperform when they impose visible ceremony on routine scholarly tasks. Effective DeSci UX, by contrast, treats verifiability and portability as *by-products* of workflows researchers already perform (e.g. submission, review, curation, replication) rather than as new steps.

A first pattern is progressive trust and identity binding. Researchers should be able to authenticate with familiar credentials (institutional SSO or ORCID) and optionally bind that identity to a DID and credential wallet behind the scenes. This supports selective disclosure and cryptographic signing for high-stakes events without imposing key management on novices. Recovery and delegation (to institutional agents or lab managers) are surfaced as ordinary settings rather than as security primitives, reducing abandonment. (ORCID, n.d.) (W3C, 2022; W3C, 2025)

A second pattern is **embedded provenance**. Rather than asking users to complete standalone metadata forms, the system emits machine-readable provenance from contextual cues: when a user deposits a dataset, the deposit action mints a PID, captures fixity, and assembles an RO-style package; when a user triggers an analysis in the platform's runner, the environment emits a signed execution record with inputs, parameters, and checksums. The user simply sees a "complete deposit" confirmation and human-readable citations; the provenance graph accrues invisibly.

A third pattern concerns **portable assessments**. Reviews and replications are authored in the same editor used for manuscripts or notebooks, then signed and bound to the relevant PIDs with one click. Templates and rubrics are embedded in the editor; conflicts of interest are declared via a short, structured interaction. This design makes the "review as object" paradigm feel like regular authoring, while enabling cross-venue portability and machine-actionable discovery.

A fourth pattern is **workflow-aware guidance**. Many frictions arise from uncertainty about what constitutes sufficient documentation, which license to select, or how to cite code and data. In high-performing systems, guidance is contextual and adaptive: the license picker suggests

defaults based on discipline and funder requirements; a checklist lights up as reproducibility conditions are satisfied (environment pinned, seed fixed, data accessible); warnings surface if a deposit conflicts with a repository's terms (e.g., personal data). This raises the floor of quality without adding cognitive load.

Finally, **integration over replacement** is essential. Researchers operate within ecosystems (editorial platforms, Git hosting, domain repositories, CRIS/RIMS). UX that interrupts these flows is punished with non-use. Instead, the DeSci layer should expose light-touch extensions: repository actions that auto-emit credentials; CI hooks that notarize releases; plugins that parse notebooks for reproducibility signals. Where replacement is unavoidable (e.g., for credential issuance), it must be justified by tangible benefits such as easier compliance or faster recognition.

4.7 Security, privacy, and compliance considerations

Security and privacy in DeSci systems must be specified with the same rigor as functional features. A credible design begins with a **threat model**: adversaries may attempt to exfiltrate sensitive data, tamper with artifacts or provenance, impersonate reviewers, launder reputation via forged credentials, or manipulate incentive programs (sybil and collusion attacks). Because the platform aspires to portability and automation, the blast radius of a flaw can be large; the system must therefore assume that components will fail and design for graceful degradation rather than perfect safety.

Confidentiality is context dependent. Much research data are not sensitive, but where personal, clinical, or proprietary data are involved, confidentiality controls must be first-class. Practical patterns include compute-to-data topologies that keep raw data in controlled enclaves; policy-based encryption with decentralized access control (keys issued against verifiable policies); and auditable access where every grant and run is logged as a verifiable event. TEEs offer hardware-backed attestation for certain workloads; they must be accompanied by supply-chain risk management and fallbacks for enclave failure. Zero-knowledge proofs or verifiable computation can provide strong guarantees for specific classes of analysis, but they are not yet general replacements for TEEs and replayable pipelines; their use should be targeted where proofs materially change trust or compliance posture. (GA4GH Cloud Work Stream, n.d.) (DataSHIELD Collaboration, n.d.) (Costan & Devadas, 2016)

Integrity requires that artifacts and their lineage resist tampering. Content addressing and fixity checks provide object-level integrity; notarization of lifecycle events (preregistration, deposit, review) hardens against backdating or silent edits. Provenance graphs should be signed and, where feasible, replicated across independent stores to avoid single points of failure. Reviewer and replication credentials require **issuer governance** (allowlists or accreditation), **revocation** mechanisms, and **double-spend** prevention for attestations that carry incentives.

Privacy extends beyond secrecy to data minimization and purpose limitation. Default flows should avoid collecting personal data unrelated to research utility; when personal data are collected (e.g., for conflicts-of-interest handling), they should be compartmentalized, encrypted at rest, and retained only as long as necessary. Selective disclosure of credential attributes reduces over-sharing. When telemetry is used to improve UX or detect abuse, it should be aggregated and, where possible, differentially private.

Compliance is a moving target spanning copyright and licensing, data protection (e.g., GDPR), research ethics, and funder mandates. The platform should encode compliance into the product: license pickers constrained by policy; machine-readable data-use conditions attached to objects; consent metadata captured at deposit; and automated checks that block obviously unlawful flows (e.g., exporting restricted data to disallowed jurisdictions). For auditability, the system should produce an evidence bundle sufficient to demonstrate due diligence to institutions and regulators. (EUR-Lex, 2016)

A final consideration is **operational resilience**. Decentralization does not eliminate operational responsibilities. Incident response (revocation of compromised keys, isolation of malicious providers, takedown of misissued credentials) must be designed *ex ante*, with published playbooks, emergency powers under time-boxed mandates, and transparent post-mortems. Treasury and incentive programs should include rate limits and anomaly detection to contain abuse while preserving usability.

5 Market Analysis

5.1 Market definition and segmentation

We define the DeSci market as the set of protocols, platforms, and communities that apply verifiability and open standards to the research lifecycle, with economic primitives that finance shared infrastructure and domain specific inquiry. This boundary includes Web3 native projects that fund or govern research agendas, decentralized data and compute infrastructures that support provenance and reproducibility, and open scholarly graphs that enable attribution and discovery. As an empirical anchor, we include projects tracked under CoinGecko's "Decentralized Science (DeSci)" category and then locate each within a functional value chain.

Table 1: Market Layer Focus Areas and Mapped Topics

Market Layer	Focus	Mapped topics (section 4)
L1. Funding BioDAOs	Funding, IP, governance	Governance, incentives, compliance and IP, user experience and adoption
L2. Data and Knowledge infrastructure	Persistent identifiers, metadata, registries, knowledge graphs	Identity, attribution, data stewardship, verifiability, interoperability
L3. Attribution and Integrators	Peer review, reputation, community curation	Attribution; user experience and adoption; governance; standards alignment
L4. Storage and Compute	Content-addressed storage, compute-to-data, decentralized compute	Data and compute verifiability; compliance; interoperability

Funding BioDAOs. A visible portion of the category comprises research funding communities that specialize by field and use tokens to coordinate grants, IP options, and translational milestones. **VitaDAO** focuses on longevity science with a portfolio model and community governance. **HairDAO** is patient owned and directs early stage funding toward hair loss research. **ValleyDAO** finances synthetic biology for climate and industrial biotechnology, with milestone based releases and flexible IP terms. **AthenaDAO** concentrates on women's health, including reproductive technologies and menopause related research. **PsyDAO** funds psychedelic science and operates grant programs that engage students and early career researchers. These initiatives exemplify a DeSci specific approach to capital formation and prioritization, and they provide structured case material for incentive design and governance analysis in subsequent sections.

Data, knowledge, and healthcare infrastructure. Several projects in the category provide data layers that are directly relevant to provenance, interoperability, and compute to data patterns. **OriginTrail** operates a decentralized knowledge graph that publishes verifiable knowledge assets and supports resolution across multiple chains, which is applicable to research metadata and discovery services. **Galeon** develops a healthcare data platform that integrates electronic health records and supports privacy preserving training of medical AI, with recent announcements about partnerships for ultrasound devices and population health analytics. These projects illustrate how DeSci infrastructures can combine content addressing, semantic schemas, and

consent aware data flows, with implications for our evaluation criteria on verifiability and compliance.

Attribution and reputation, and integrators. Tokens also appear as incentive layers for scholarly contribution. **ResearchCoin (RSC)** underpins the ResearchHub platform, and is used to reward curation, discussion, and open peer input, which maps to our identity and attribution segment and to portable assessment patterns. We treat scholarly knowledge graphs and PID infrastructures as integrators in this market, since they enable cross platform discovery and linking. **OpenAlex** is a prominent example with a free API and daily updates that index research outputs, authors, venues, and institutions. Together, these elements ground the identity and interoperability signals in our rubric.

Compute and marketplaces, and emerging hybrids. A subset of listed assets reflects compute marketplaces and hybrid stacks that are adjacent to DeSci but relevant when workflows require elastic resources or verifiable execution. While our comparative scoring emphasizes research directed designs, we include such projects in scanning to the extent that they support content addressed pipelines, attestable runs, or compute to data orchestration in research settings. The category also includes **Bio Protocol (BIO)**, a curation and funding meta layer that incubates BioDAOs and introduces tokenized IP mechanisms, which is pertinent to our analysis of governance patterns and sustainability of research pipelines.

This segmentation provides a concrete roster of exemplars for the comparative analysis. It also clarifies that “DeSci” is not a single vertical, rather a set of interacting layers that combine funding and governance, identity and attribution, provenance aware data infrastructure, storage and compute assurances, and integrative knowledge graphs. These layers are the basis for the evaluation criteria and the provider profiling that follow.

5.2 Comparative framework

Comparing heterogeneous DeSci platforms, protocols, and tooling requires a common evaluative lens that is both principled and operational. In this report we adopt a multi-criteria framework that synthesizes architectural, governance, interoperability, and adoption considerations into a single score while preserving the underlying profile of strengths and weaknesses.

The aim is not merely to rank vendors, but to surface trade-offs that inform requirements (D18), guide wireframing priorities (WP4), and constrain the technical specifications and roadmap (WP5). To that end, the rubric is grounded in the conceptual foundations set out in Sections 3, 4 and parameterized with weights reflecting the project’s TRL-1 design emphasis.

The framework rests on six criteria, each capturing a distinct but interdependent facet of decentralized research collaboration. **Governance & decentralization** assesses decision rights, upgrade processes, credible neutrality, and resilience; features that shape long-run evolvability and trust. **Identity/attribution & interoperability** focuses on alignment with scholarly identifiers

(e.g., ORCID, CRediT) and Web3 identity primitives (DID/VC), together with schema and API interoperability that enable cross-system credit and verification. **Data/compute & verifiability** evaluates adherence to FAIR principles, provenance registries, and the extent to which storage and compute are verifiable and reproducible. **Incentives/tokenomics & sustainability** examines the robustness of economic mechanisms (e.g., reviewer incentives, public-goods funding, retroactive funding) and fiscal durability. **UX/onboarding & adoption** captures the practical frictions researchers face, such as documentation, integrations, and observed traction, while **Compliance/ethics & IP posture** addresses rights management, GDPR/data sovereignty, and responsible research standards. Detailed rating anchors and evidence expectations for each criterion are provided in **Appendix C**.

Each criterion is rated on a five-point ordinal scale using the anchors in Appendix C. Ratings are aggregated via a weighted linear composite:

$$S = \sum_{i=1}^{6} w_i r_i with \sum_i w_i = 1,\ r_i \in \{1, \ldots, 5\}.$$

For the **TRL-1 design phase**, the adopted weights are: Governance & decentralization **0.20**; Identity/attribution & interoperability **0.18**; Data/compute & verifiability **0.22**; Incentives/tokenomics & sustainability **0.12**; UX/onboarding & adoption **0.18**; Compliance/ethics & IP posture **0.10** (see Appendix D).

Weights privilege criteria that are most architecture-shaping at TRL-1. In particular, **Data/compute & verifiability** and **Governance** receive the highest weights because they materially constrain the design space and determine the credibility of scientific workflows over time. **Identity/attribution & interoperability** and **UX/onboarding** are weighted next, reflecting their direct impact on researcher adoption and cross-ecosystem integration. **Incentives/tokenomics** and **Compliance/ethics** remain essential, but at this stage they are treated as design constraints rather than primary differentiators; their relative importance may increase in later piloting phases as real-world usage and regulatory exposure deepen.

Ratings must be evidence-led. Acceptable evidence includes maintained documentation (with versions/dates), published schemas and APIs, standards-conformance statements, on-chain governance records, audit reports, integration catalogs, adoption metrics, and independent evaluations. Where evidence is ambiguous or incomplete, we apply a conservatism rule and award the lower of the plausible ratings, logging the uncertainty and required follow-ups. Missing data default to a provisional 2 (“limited”) with a corrective action noted in the evidence log. To reduce subjectivity, two assessors rate each provider independently; disagreements of ≥1 point on any criterion are reconciled via a brief written memo that cites evidence and records the decision. All sources, versions, and access dates are recorded in the per-provider evidence sheet (Appendix D).

The composite score is interpreted as a decision aid rather than a mechanical gate. As guidance, scores at or above 4 indicate strong alignment with our design objectives and merit inclusion in

WP4 scenarios; scores in the 3, 4 range are viable but may require mitigation plans or targeted design accommodations; anything below 3 indicates a poor fit.

5.3 Landscape overview

Funding communities and BioDAOs

Funding communities specialize by field and couple community governance to grant making and translational milestones. The segment includes five core exemplars and two additional providers to improve coverage. VitaDAO focuses on longevity science with a portfolio of early-stage projects and fellowships. HairDAO is a patient centered network that supports research into hair loss and related interventions. ValleyDAO concentrates on synthetic biology for climate and industrial biotechnology, with an explicit non-clinical scope. AthenaDAO addresses women's health, including reproductive technologies and menopause. PsyDAO supports psychedelic science through community grants. Bio Protocol is a meta-layer for BioDAOs that curates projects and introduces tokenized impact mechanisms, which is relevant to incentive design. LabDAO coordinates open wet lab access and community tooling, which illustrates early program structures outside a single disease area. Across these initiatives the observable variables include proposal pipelines, decision latency, participation, transparency of treasuries, and production of public artifacts.

Data and knowledge infrastructure

This segment provides the provenance-aware data layers and semantic scaffolding needed for verifiable research objects. OriginTrail operates a decentralized knowledge graph for verifiable knowledge assets that can be resolved across chains. Galeon develops a healthcare data platform that integrates electronic health records and supports privacy preserving model training and compute to data concepts. Open Research Knowledge Graph, also referred to as ORKG, enables machine actionable scholarly knowledge and structured comparisons with public APIs. To broaden coverage, we recognize OpenMined as a community that builds privacy preserving machine-learning tooling, particularly PySyft, which is used in controlled data contexts and relates to compute to data patterns. We also highlight Zenodo as a reference integrator for repository practice, since it demonstrates DOI versioning, ORCID integration, and reproducible packaging patterns that DeSci projects can emulate or interoperate with. Together these exemplars show how consent aware schemas, provenance models, and content addressed references can be joined into usable knowledge services.

Attribution, review, and integrators

Attribution and integrators connect identities, outputs, venues, and evaluation signals. OpenAlex provides a high coverage scholarly graph with a free API that indexes works, authors, venues, institutions, and concepts. Crossref and DataCite deliver persistent identifiers and rich metadata for publications and datasets, which serve as anchors for interoperability. ResearchHub, together with ResearchCoin, aligns incentives for contribution and public curation. To strengthen coverage on open evaluation, we add OpenReview, which offers configurable open peer review workflows and public APIs, and PREreview, which supports community reviews of preprints and reviewer training. We also include Sciety, which aggregates evaluations of preprints across multiple communities, and PubPub, which provides open infrastructure for community run journals and conferences. This cluster demonstrates how portable contribution records, identifier alignment, and open interfaces can enable recognition and discovery across platforms.

Storage and compute substrate

The substrate combines verifiable transport and durable persistence, together with orchestration patterns that respect data governance. IPFS offers content addressed transport and verifiable retrieval that many systems already use for artifact integrity. Filecoin introduces market backed durability based on proofs of replication and spacetime, while Arweave offers an endowment style persistence model that suits archival use cases. Ocean Protocol brings compute to data orchestration for governed datasets, returning results and logs without releasing raw data. Bacalhau schedules container and WebAssembly jobs near content addressed inputs, which fits compute over data scenarios. Akash Network provides decentralized compute resources for batch and artificial intelligence workloads. To round out the view we add Golem, a marketplace for distributed compute cycles with a mature developer ecosystem, and iExec, which focuses on confidential computing and verifiable off chain computation using trusted hardware. Across these providers, the key design variables are fixity and PID binding, durable storage policies, execution provenance that end users can verify, and simple paths for integrators to adopt these assurances.

5.4 Adoption signals and traction metrics

Adoption and traction are uneven across segments, therefore we treat them as multi dimensional. We report descriptive metrics where reliable public sources exist, and otherwise use standardized signals that are comparable across projects.

Funding communities and BioDAOs. Primary signals include number of proposals submitted and funded, median time from proposal to decision, grant volume by quarter, proportion of follow on financing, and governance participation rates for budget and policy votes. A secondary layer includes contributor retention, the share of successful deliverables that proceed to publications or company formation, and evidence of open licensing where appropriate. These are extractable

from public governance portals and grant trackers where available, for example VitaDAO's project pages and similar sources for other BioDAOs.

Data and knowledge infrastructure. Signals include documented integrations with repositories or hospital systems, stability and coverage of public APIs, versioned specifications, and external adoption events such as conferences and third party deployments. OriginTrail's public documentation of a decentralized knowledge graph and Galeon's recent device to EHR announcements are examples of evidentiary signals.

Attribution, review, and integrators. Signals include ORCID and DOI write backs, volume of public reviews and curations, growth in active contributors, API call volumes, and data freshness in knowledge graphs. OpenAlex publishes a transparent update cadence and API, and projects such as ResearchHub document token mediated incentives and community programs.

Normalization and caveats. We normalize participation rates by the eligible voter base to avoid inflated turnout. We weight grants by stage and scope to avoid treating micro awards and multi year programs as equivalent. We record uncertainties when the only sources are marketing materials or social media. Where numbers are not auditable, we prefer qualitative categories with evidence citations rather than speculative figures.

Table 2: Funding communities and BioDAOs

Metric	Operational definition	Source (URL)	Collection method	Normalization / notes
Grants funded per quarter	Number of grants approved in the quarter, with amounts and short descriptions	https://www.vitadao.com/researchers	Manual scrape or program API if available; record decision date and amount	Normalize by treasury size for comparability; list currency and denomination policy
Proposal decision latency	Median days between proposal submission and recorded decision	https://www.vitadao.com	Collect timestamps from proposal and vote records; compute per quarter	Report median and interquartile range; exclude withdrawn proposals with note
Voter participation rate	Turnout as share of eligible voters for budget or policy votes	https://www.vitadao.com	Count eligible and participating voters per vote; aggregate quarterly	Report mean and standard deviation; annotate changes in eligibility rules
Outputs ratio	Share of funded projects that produce public artifacts, for example preprints, datasets, or code, within 12 months	Program pages across providers	Follow grant update links; verify DOIs or repository links	Denominator is grants older than 12 months; list artifact types collected

Table 3: Data and knowledge infrastructure

Metric	Operational definition	Source (URL)	Collection method	Normalization / notes

API coverage and stability	Number of documented endpoints, versioning scheme, and presence of rate limits and status page	https://docs.origintrail.io	Manual review of docs; log endpoint families, version headers, rate limits	Record doc version and access date; capture ‘breaking change’ policy
Documented integrations	Count of named repositories, EHRs, or third party systems with public documentation	https://www.galeon.care	Manual review of integration and partner pages	Include only named, verifiable integrations; separate pilots from production
Consent and data use metadata	Presence of consent or data use fields in published schemas or SDKs	https://www.galeon.care	Inspect schema references and SDK examples	Record schema version; note whether fields are required or optional
Compute-to-data evidence	Presence of documented patterns or case studies where algorithms run without exposing raw data	https://www.galeon.care	Manual review of whitepapers and case pages; capture workflow diagrams	Tag sector, for example clinical or social data; note audit and logging features

Table 4: Attribution, review, and integrators

Metric	Operational definition	Source (URL)	Collection method	Normalization / notes
Identifier write backs	Presence and scope of ORCID and DOI write backs or exports	https://docs.openalex.org	Manual review of API docs and integration guides	Qualitative scale: none, export only, bidirectional write back; note rate limits
API freshness and volume	Declared update cadence and current entity counts (works, authors, venues, institutions)	https://docs.openalex.org	Record stated cadence; retrieve counts from status or docs	Capture date; flag when cadence deviates from declared policy
Public curation or review volume	Number of public reviews, evaluations, or curated items per month	https://docs.researchhub.com	Manual count from platform pages or APIs; sample over three months	Report average per month; annotate moderation policies if available
Coverage of PIDs in records	Share of records with ORCID, DOI, or ROR identifiers present	https://docs.openalex.org	Sample one thousand records per entity type via API; compute shares	Stratify by subfield if feasible; report confidence intervals

Table 5: Storage and compute substrate

Metric	Operational definition	Source (URL)	Collection method	Normalization / notes
PID to CID binding coverage	Share of artifacts where a persistent identifier is linked to a content identifier and fixity information	https://docs.ipfs.tech	Sample artifacts from repositories that claim IPFS support; check landing pages or manifests	Report per repository and overall; note checksum algorithms used

Durability commitments	Presence of renewal or endowment policies that specify minimum retention periods	https://filecoin.io/filecoin.pdf	Manual review of protocol docs and provider policies	Categorize as renewal required, endowment, or ad hoc; capture minimum period
Execution provenance availability	Presence of signed job records, logs, or attestations visible to end users	https://www.bacalhau.org	Manual review of job metadata and docs; attempt a sample run where feasible	Note signature or attestation mechanism; record verification steps
Retrieval performance proxy	Median time to first byte for retrieval from public gateways or provider endpoints	https://docs.ipfs.tech	Measure with standardized requests from two regions; record five trials per object	Report medians and variability; annotate gateway or endpoint names

5.5 Risks, constraints, and externalities

DeSci initiatives operate in environments that are technically ambitious and institutionally sensitive. We therefore assess risks and constraints explicitly, and we record the externalities that affect adjacent communities.

Regulatory and compliance risks. Health and social data often fall under strict regimes. Projects that collect or process such data must demonstrate compliance with the General Data Protection Regulation and with sector specific rules, and must separate personally identifiable information from public artifacts. Compute to data patterns and auditable access logs are useful mitigations but do not remove obligations. Announcements of device to EHR integrations illustrate both the opportunity and the compliance burden.

Governance and incentive risks. Token weighted governance can concentrate power. Low participation can undermine legitimacy. Incentive programs can be gamed when rewards are not tied to verifiable outcomes. We therefore treat governance participation rates, upgrade key disclosures, treasury transparency, and the presence of appeal routes as gating conditions in scoring.

Operational and sustainability risks. Storage persistence depends on active pinning or on paid commitments. Documentation and API stability are essential for integrators. BioDAO treasuries are exposed to market volatility, which can disrupt research funding cycles. We therefore examine treasury runway and diversification, and we record whether grants are denominated in a stable unit of account.

Reputational and scientific quality risks. Hype cycles and meme assets can erode trust in serious work. We distinguish research directed projects from speculative categories and we require evidence of methods, evaluation, or standards adoption. The CoinGecko category is useful for discovery, although it includes entertainment assets that we do not analyze as research infrastructure.

Externalities and ecosystem effects. Some initiatives create positive spillovers for open science even when they are not strictly decentralized, for example by improving provenance capture or by expanding access to replication funding. Others may impose costs on repositories and indexers if they rely on undocumented scraping or unstable APIs. Our recommendations in this report prioritize alignment with common standards to maximize positive externalities and to reduce integration burdens.

6 Comparative Analysis & Findings

This section synthesizes the literature review of section 4 with the market segments of section 5 to support an integrated analysis. We adopt the following segmentation matrix.

Table 6: Literature–Market Segmentation Matrix

Section 5: Market Analysis		Section 4: Literature	
Market segment	**Examples**	**Topics aligned**	**Core standards or anchors**
Funding communities and BioDAOs	VitaDAO, HairDAO, ValleyDAO, AthenaDAO, PsyDAO	Governance and funding, UX and adoption, Compliance and IP	Open governance records and policies, treasury disclosures, contribution credentials mapped to CRediT where applicable
Data and knowledge infrastructure	OriginTrail, Galeon	Data management and provenance, Compliance and IP, Identity and attribution	PROV, RO-Crate, data use and consent metadata, DID or VC for access and audit trails
Attribution and integrators	ResearchCoin, OpenAlex	Identity and attribution, Interoperability, UX and adoption	ORCID, DOI metadata from Crossref or DataCite, DID or VC for portable contribution records, documented APIs
Storage and compute substrate	IPFS, Filecoin, Arweave	Storage and compute, Data verifiability	Content addressing and fixity, economic durability or endowment persistence, reproducible or attestable execution

6.1 SWOT analysis by segment

Segment 1: Funding communities and BioDAOs

Strengths. Visible pipeline of field-specific grants, community governance, transparent discussions and decisions. Thematic focus and community mobilization.

Weaknesses. Participation and voter turnout variance, treasury volatility, heterogeneous grant reporting, limited integration with scholarly identifiers.

Opportunities. Issue portable contribution credentials, link to ORCID and DOIs, adopt outcome-based disbursements similar to retroactive funding, publish programmatic APIs for grant data.

Threats. Regulatory uncertainty for tokens, reputational risks from non-research assets listed in the same category feeds.

Segment 2: Data and knowledge infrastructure

Strengths. Provenance friendly models, emphasis on verifiable knowledge assets and governed health data flows. OriginTrail publishes a decentralized knowledge graph, Galeon documents compute to data concepts in health records.

Weaknesses. Sector specific integration burdens, especially for clinical data, uneven publication of schemas and consent metadata.

Opportunities. Standardize on PROV and RO-Crate for portability, expose crosswalks to Crossref and DataCite metadata, adopt DID or VC for access grants and audit trails.

Threats. Compliance and data protection risks, vendor lock-in via undocumented integrations.

Segment 3: Attribution and integrators

Strengths. Clear identity and metadata anchors, with OpenAlex offering a free, high coverage API, and ResearchCoin aligning incentives for contribution and review.

Weaknesses. Fragmented credential portability, limited write-backs to ORCID or PID registries from Web3-native platforms.

Opportunities. Bind reviewer and curator activity to verifiable credentials that resolve to ORCID, publish export and import utilities, integrate with OpenAlex concepts and works endpoints.

Threats. Low-signal incentives can attract gaming, burdens on moderators without clear governance.

Segment 4: Storage and compute substrate

Strengths. Secure storage with no single point of failure, economic or endowment-based durability. IPFS, Filecoin, and Arweave offer distinct persistence models that can back PID-linked artifacts.

Weaknesses. Retrieval latency variance, dependency on pinning or payment policies, limited turnkey reproducibility signals for non-expert users.

Opportunities. Default PID plus fixity binding, policy backed pinning, verifiable execution records bundled with outputs.

Threats. Cost volatility for storage markets, misunderstanding of permanence claims, operational risks around gateway concentration.

6.2 Unmet needs

Funding communities and BioDAOs

1. Portable grant and contribution credentials that resolve to ORCID, with links to DOIs for funded outputs.

2. Standard program metadata and APIs for proposals, decisions, milestones, and outcomes, to enable discovery and evaluation.
3. Outcome-based disbursement modules that release funds against verifiable artifacts, for example replication credentials.

Data and knowledge infrastructure

1. PROV and RO-Crate default packaging for all data products and AI training sets, with explicit consent and data use fields.
2. DID or VC based access grants and audit events, for compute to data workflows in clinical settings.
3. Crosswalks to Crossref or DataCite schemas, and resolvable links into OpenAlex to improve discovery.

Attribution and integrators

1. Reviewer and curator credentials that are verifiable, selectively disclosable, and linked to PIDs, to travel across venues.
2. Write-back and export utilities to ORCID and DOI registries, with changelogs and versioned APIs.
3. Topic and concept alignment using OpenAlex endpoints to improve findability and analytics.

Storage and compute substrate

1. Turnkey “PID plus fixity” packaging with human readable citations for IPFS or Filecoin backed artifacts.
2. Durable storage policies, for example paid renewal or endowment commitments, with visible status signals.
3. Standard execution provenance records, emitted by runners, that can be verified without specialist knowledge.

7 Key insights and implications

This section enumerates twelve evidence-backed insights derived from the literature synthesis in section 4 and the market analysis in section 5. Each insight states a short rationale and a concrete implication for WP4 and WP5. The intent is to provide clear, testable guidance for stakeholder validation in D18 and for design and specification work in subsequent work packages.

7.1 Identity-first, DID optional

Rationale. ORCID is the canonical person identifier across the scholarly graph and is already integrated with repositories and knowledge indexes. Decentralized identifiers and verifiable credentials can be layered to express portable claims when needed.

Implication for WP4 and WP5. Implement ORCID sign-in as the default, allow optional binding to a decentralized identifier, and support selective disclosure of verifiable credentials for review, replication, and stewardship roles.

7.2 Provenance by default

Rationale. Provenance is most reliable when it is emitted automatically during deposit and execution, not when it is added retrospectively. FAIR aligned metadata, PROV based lineage, and RO Crate packaging enable machine actionability.

Implication for WP4 and WP5. Instrument repository deposits and compute runners to emit provenance and package research objects automatically, then expose both human-readable citations and machine-readable graphs.

7.3 Content-addressed storage with durable commitments

Rationale. Content addressing provides integrity and deduplication, while market or endowment backed storage provides durability over time.

Implication for WP4 and WP5. Bind persistent identifiers to content identifiers, publish pinning and renewal policies, and surface storage status to users in a transparent way.

7.4 Compute to data for governed datasets

Rationale. Moving computation to the data reduces exposure of sensitive information and aligns with common data protection obligations in health and social research.

Implication for WP4 and WP5. Offer compute to data orchestration, record grants and runs as verifiable events, and minimize retention of personal data by default.

7.5 Portable assessment objects

Rationale. Reviews, replication attempts, and curation events can be modeled as first class objects that travel across venues and time, which supports recognition and reuse.

Implication for WP4 and WP5. Support reviewer and replicator credentials, provide export and write back utilities that connect to ORCID and repository records.

7.6 Incentives tied to verifiable outcomes

Rationale. Funding and bounty programs are most effective when rewards are conditioned on artifacts that are resolvable and verifiable, for example reproducible runs or linked datasets.

Implication for WP4 and WP5. Prototype outcome-based disbursement modules that release funds when specified verifiable artifacts are present and linked to their provenance records.

7.7 User experience and integrations

Rationale. Researchers adopt tools that integrate with their existing workflows and that reduce cognitive load. Cryptographic ceremony should be hidden behind familiar flows.

Implication for WP4 and WP5. Prioritize integrations with repositories, code hosting platforms, editorial and review tools, and provide simple account recovery and key management experiences.

7.8 Governance transparency

Rationale. Public roadmaps, disclosure of upgrade keys and policies, participation metrics, and dispute or appeal routes are baseline signals of credible neutrality.

Implication for WP4 and WP5. Publish governance documents and metrics, define delegated roles with clear remits, adopt time-boxed emergency powers with post-mortems, and document treasury policies.

7.9 Open APIs and schemas

Rationale. Stable, versioned APIs with documented schemas and changelogs reduce integration risk and enable ecosystem reuse.

Implication for WP4 and WP5. Publish OpenAPI specifications, provide version headers and deprecation policies, and maintain changelogs and usage examples for developers.

7.10 Segment-specific adoption signals

Rationale. A single metric cannot represent adoption across funding communities, data platforms, integrators, and storage or compute layers.

Implication for WP4 and WP5. Track a small set of segment-specific indicators, for example grants per quarter and decision latency for funding programs, API coverage and freshness for data platforms, identifier write backs for integrators, and fixity plus durability signals for storage.

7.11 Compliance and privacy by design

Rationale. Legal and ethical obligations set hard constraints on architecture, retention, and auditability, especially for personal and clinical data.

Implication for WP4 and WP5. Minimize personal data collection, separate contact information from research data, embed consent and data use metadata in packages, and log access events and compute grants.

7.12 Maintain current rubric weights for the first design cycle

Rationale. Initial scoring suggests ranks are driven primarily by verifiability and identity signals, and minor weight shifts are unlikely to change design priorities at this stage.

Implication for WP4 and WP5. Retain the current weights through the first WP4 and WP5 milestones, then revisit after stakeholder feedback and post-project.

8 Conclusion

This report has established a defensible evidence base for decentralized research collaboration. The literature converges on standards that enable portability and trust, and the market offers concrete building blocks across funding, identity and attribution, provenance aware data infrastructure, and storage and compute. We have found that the main risks are not technical, but relate to governance quality, usability, and compliance. The most promising path is to bind person and object identifiers to verifiable claims, to capture provenance automatically, and to expose these assurances through interfaces that feel familiar to researchers.

Key Insights Summary

This deliverable (D17: *Literature Review and Market Analysis Report*) was completed in line with the project plan. It identifies **over ten key insights** from the literature and market analysis that guide the design of the DeSciCoLab platform and inform subsequent work packages:

1. Decentralized identity systems combining ORCID and DIDs are essential for verifiable attribution.
2. Provenance must be captured automatically within normal research workflows.
3. FAIR data practices remain inconsistent, highlighting a need for interoperable standards.
4. Hybrid storage solutions balance verifiability, scalability, and cost.
5. Compute verifiability remains limited; replayable pipelines are the most practical baseline.
6. Governance models should mix off-chain deliberation with on-chain transparency.
7. Effective incentives link rewards to measurable, verified outcomes.
8. UX should make verification and compliance seamless for users.
9. Legal and ethical safeguards must be embedded in metadata and workflows.
10. The DeSci ecosystem remains fragmented, with gaps in interoperability and governance maturity.
11. Verifiability should emerge naturally from everyday research activity.

Together, these insights form a clear evidence base for the **wireframe development (WP4)** and **technical specifications and roadmap (WP5)** phases.

Two practical steps follow. First, we will validate the priorities and adoption barriers with stakeholders in D18, using the survey and interview instruments keyed to the six criteria. Second, we will translate the unmet needs into interaction patterns and technical components in WP4 and WP5, beginning with identity binding, provenance by default, portable assessments, and

durable storage policies. With these steps, the project can deliver a platform that reduces coordination costs, improves attribution and reproducibility, and meets the legal and ethical standards required for real world research.

9 References & Bibliography

Akash Network. (2025). Akash Network documentation. https://akash.network/docs

Anagen by HairDAO. (2025). Anagen. https://anagen.xyz
Arweave. Williams, S., et al. (2018/2023). Arweave: Permanent Information Storage Protocol. https://www.arweave.org

AthenaDAO. (2025). AthenaDAO. https://www.athenadao.co

Bacalhau Project. (2025). Bacalhau. https://www.bacalhau.org

Benet, J. (2014). IPFS: Content Addressed, Versioned, P2P File System. https://arxiv.org/abs/1407.3561

Brand, A., Allen, L., Altman, M., Hlava, M., & Scott, J. (2015). Beyond authorship: Attribution, contribution, collaboration, and credit. Learned Publishing, 28(2). https://doi.org/10.1087/20150211

Buterin, V., Hitzig, Z., & Weyl, E. G. (2018). Liberal Radicalism: A Flexible Design for Philanthropic Matching Funds. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3243656

Crossref. (2025). Crossref. https://www.crossref.org

DataCite. (2025). DataCite. https://datacite.org

DataSHIELD Collaboration. (n.d.). DataSHIELD: Privacy-preserving federated analysis. https://www.datashield.org

eLife. (2024). eLife's reviewed-preprint model. https://elifesciences.org

EUR-Lex. (2016). General Data Protection Regulation (Regulation (EU) 2016/679). https://eur-lex.europa.eu/eli/reg/2016/679/oj

Filecoin. (2025). Filecoin documentation. https://docs.filecoin.io

Filecoin. Protocol Labs. (2017). Filecoin: A Decentralized Storage Network. https://filecoin.io/filecoin.pdf

GA4GH Cloud Work Stream. (n.d.). Bringing compute to the data: Standards and implementations. https://www.ga4gh.org/workstreams/cloud

Galeon. (2025). Galeon. https://www.galeon.care

Golem Network. (2025). Golem documentation. https://docs.golem.network

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? Science Translational Medicine, 8(341). https://doi.org/10.1126/scitranslmed.aaf5027

HairDAO. (2025). HairDAO. https://www.hairdao.xyz

Hypercerts Foundation. (2023). Hypercerts: A data layer for impact funding. https://hypercerts.org

IPFS. (2025). IPFS documentation. https://docs.ipfs.tech

NISO. (2025). CRediT—Contributor Roles Taxonomy. https://credit.niso.org

Ocean Protocol. (2024). Compute-to-Data documentation. https://docs.oceanprotocol.com

OpenAlex. (2025). OpenAlex. https://openalex.org

OpenReview. (2025). OpenReview. https://openreview.net

Optimism Collective. (n.d.). Retroactive Public Goods Funding. https://community.optimism.io/docs/governance/retroPGF/

ORCID. (n.d.). Connecting research and researchers. https://orcid.org

OriginTrail. (2025). OriginTrail documentation. https://docs.origintrail.io

PREreview. (2025). PREreview. https://prereview.org

PubPub. (2025). PubPub platform. https://www.pubpub.org

ResearchHub. (2025). ResearchHub. https://www.researchhub.com

ResearchHub Foundation. (2025). ResearchCoin. https://docs.researchhub.com

Research Object Community. (2025). RO-Crate specification. https://www.researchobject.org/ro-crate

Review Commons. (n.d.). Portable, journal-agnostic peer review. https://www.reviewcommons.org

Sciety. (2025). Sciety. https://sciety.org

Soiland-Reyes, S., Sefton, P., Crosas, M., Castro, L. J., et al. (2022). Packaging research artifacts with RO-Crate. Data Science, 5(2). https://doi.org/10.3233/DS-210053

ValleyDAO. (2025). ValleyDAO. https://www.valleydao.bio

VitaDAO. (2025). VitaDAO. https://www.vitadao.com

W3C. (2013). PROV—Provenance data model. https://www.w3.org/TR/prov-overview

W3C. (2013). PROV-O: The PROV Ontology. https://www.w3.org/TR/prov-o/

W3C. (2022). Decentralized Identifiers (DIDs) v1.0. https://www.w3.org/TR/did-core/

W3C. (2025). Verifiable Credentials Data Model v2.0. https://www.w3.org/TR/vc-data-model-2.0/

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18

10 Appendix A. Literature review search strings

Protocol notes. Time window is 2016 to present, extended backward only when a pre-2016 work is clearly seminal. Language is English. Wherever possible, searches are restricted to title, abstract, and keywords for precision. For each source we record the final query string, filters, the date executed, and the export file name. Duplicates are removed before screening. The core concept clusters are:
(1) decentralized science, open science with blockchain, research collaboration with blockchain or distributed ledger,
(2) identity and attribution, verifiable credentials, decentralized identifiers, ORCID, CRediT,
(3) provenance and FAIR, reproducibility, research objects,
(4) governance and incentive mechanisms, quadratic funding, retroactive funding,
(5) decentralized storage and compute, verifiable compute, IPFS, Filecoin, Arweave.

Scopus (TITLE ABS KEY, 2016 to present, English):
("decentralized science" OR "decentralised science" OR DeSci OR ("open science" W/3 blockchain) OR ("research collaboration" W/3 (blockchain OR "distributed ledger"))) AND (identity OR attribution OR "verifiable credential*" OR "decentralized identifier*" OR DID OR provenance OR FAIR OR reproducib* OR governance OR DAO OR "quadratic funding" OR "retroactive funding" OR "verifiable comput*" OR "decentralized stor*" OR IPFS OR Filecoin OR Arweave)

Web of Science (Topic, 2016 to present, English):
("decentralized science" OR "decentralised science" OR DeSci OR ("open science" NEAR/3 blockchain) OR ("research collaboration" NEAR/3 (blockchain OR "distributed ledger"))) AND (identity OR attribution OR "verifiable credential*" OR "decentralized identifier*" OR DID OR provenance OR FAIR OR reproducib* OR governance OR DAO OR "quadratic funding" OR "retroactive funding" OR "verifiable comput*" OR "decentralized stor*" OR IPFS OR Filecoin OR Arweave)

IEEE Xplore (All Metadata, 2016 to present, English):
(("blockchain" OR "distributed ledger") AND ("open science" OR "research data" OR reproducib* OR provenance) AND (identity OR "verifiable credential*" OR "decentralized identifier*" OR FAIR OR "decentralized storage" OR "verifiable compute")) (2016, present).

ACM Digital Library (Abstract/Title/Keywords, 2016 to present, English):
("decentralized science" OR DeSci OR "open science" AND blockchain) AND (identity OR DID OR "verifiable credential*" OR provenance OR FAIR OR "decentralized storage" OR "verifiable compute") (2016, present).

arXiv (Computer science categories cs.DL, cs.CR, cs.DC, 2016 to present, English)

("decentralized science" OR DeSci OR ("open science" AND blockchain))

AND

(identity OR DID OR "verifiable credential*" OR provenance OR FAIR OR "verifiable compute" OR "decentralized storage")

SSRN (2016 to present, English)

("decentralized science" OR DeSci OR ("open science" AND blockchain))

AND

(governance OR DAO OR "quadratic funding" OR "retroactive funding" OR "public goods")

OSF (2016 to present, English)

("decentralized science" OR DeSci)

AND

(provenance OR FAIR OR "research object" OR reproducib*)

Google Scholar (curated first 200 results, 2016 to present, English)

"decentralized science" OR DeSci OR ("open science" blockchain)

11 Appendix B. Data extraction codebook and samples

Purpose. The codebook standardizes what we extract from each included source so that synthesis and comparative analysis remain auditable, repeatable, and reusable in D18, WP4, and WP5. The unit of analysis is a single source. For market profiling, the unit of analysis is a provider or protocol page.

Field list, definitions, and examples

1. **Record identifier**
   Definition. A unique key for the record in the screening log.
   Allowed values: Free text.
   Example. SCOPUS 041.

2. **Bibliographic metadata**
   Definition. Author or authors, year, title, venue, DOI or URL.
   Allowed values: Free text.
   Example. Wilkinson et al., 2016, The FAIR Guiding Principles, Scientific Data, DOI.

3. **Source type**
   Definition. Empirical, conceptual, review, technical standard, white paper.
   Allowed values: One of the five types.
   Example. Technical standard.

4. **Domain and context**
   Definition. Discipline, research stage, deployment context, and whether the work concerns human participant data.
   Allowed values: Free text with controlled tags, for example life sciences, preprint, repository deposit, human data yes or no.
   Example. Engineering, code release, repository deposit, human data no.

5. **Constructs and mechanisms**
   Definition. Core ideas and mechanisms addressed, for example identity and attribution, provenance, storage, compute, governance, incentives, user experience.
   Allowed values: Multi select from the list above, plus free text where necessary.
   Example. Identity and attribution, provenance, user experience.

6. **Methods and evidence**
   Definition. Study design, dataset, evaluation metrics, or method of technical validation, including threat model if present.
   Allowed values: Free text with tags such as experiment, case study, benchmark, standardization, specification, audit.
   Example. Case study with repository logs and replication records.

7. **Outcomes and findings**
   Definition. Key results, adoption signals, performance indicators, and reported

limitations.
Allowed values: Free text.
Example. Reduced time to deposit, increased dataset reuse, limitation in cross repository interoperability.

8. **Identity and interoperability details**
   Definition. Use of ORCID, DOIs, DIDs, verifiable credentials, schemas, and APIs.
   Allowed values: Free text, plus yes or no tags for ORCID, DOI, DID, VC.
   Example. ORCID yes, DOI yes, DID yes, VC yes, REST API with OpenAPI schema.

9. **Provenance and packaging details**
   Definition. Use of FAIR, PROV, RO Crate or equivalent, fixity, content addressing, and versioning.
   Allowed values: Free text, plus yes or no tags for FAIR, PROV, RO Crate, fixity, content addressing, versioning.
   Example. FAIR yes, PROV yes, RO Crate yes, fixity and versioning present.

10. **Storage and compute details**
    Definition. Storage model and guarantees, compute model and assurances, and whether compute to data is supported.
    Allowed values: Free text, plus yes or no tags for content addressing, market backed durability, TEE attestation, ZK proof evidence, compute to data.
    Example. Content addressing yes, market backed durability yes, compute to data yes.

11. **Governance and incentive details**
    Definition. Governance arrangements, decision rights, transparency features, and incentive mechanisms such as bounties, quadratic funding, or retroactive funding.
    Allowed values: Free text, plus yes or no tags for open governance, public treasury, quadratic funding, retroactive funding.
    Example. Open governance with public treasury, retroactive funding pilot.

12. **UX and adoption details**
    Definition. Onboarding pattern, documentation quality, integrations, and measured or reported adoption.
    Allowed values: Free text, plus one to five rating for perceived usability if the source contains a user study.
    Example. ORCID sign in, repository integration, documentation rated high by users.

13. **Compliance and IP details**
    Definition. Licensing, GDPR stance, consent metadata, data processing agreements, and data use conditions.
    Allowed values: Free text, plus yes or no tags for license stated, GDPR controls, consent metadata.
    Example. CC BY 4.0, GDPR controls present, consent metadata captured.

14. **Quality flags**
    Definition. Notes on conceptual clarity, replicability, dataset adequacy, and overall evidential strength.
    Allowed values: Free text plus one of low, moderate, high.
    Example. Conceptually clear, moderate evidence, dataset limited to one repository.
15. **Design implications tag**
    Definition. Short tag that links the finding to a requirement theme, for example identity, provenance, storage, compute, governance, incentive, UX, compliance.
    Allowed values: One or more tags from the list.
    Example. Provenance, UX.
16. **Citation string**
    Definition. The formatted citation used in the References section.
    Allowed values: APA style with URL or DOI.
    Example. Wilkinson, M. D., and colleagues, 2016, Scientific Data, DOI.

Coding rules

- For missing data, record unknown and add a follow up note in the evidence log. Do not infer capabilities that are not documented.
- For market documents and specifications, record the version and access date.
- When multiple versions of the same item exist, code the most recent stable version and note any changes that affect findings.

Examples

- A conceptual paper that proposes DID linked review credentials would be coded as type conceptual, constructs identity and attribution and user experience, identity and interoperability details with DID and verifiable credentials yes, and a design implications tag of identity.
- A repository case study that demonstrates RO Crate packaging and PROV based lineage would be coded as type empirical, constructs provenance and packaging, provenance and packaging details with FAIR yes, PROV yes, RO Crate yes, and a design implications tag of provenance.

12 Appendix C. Market scoring rubric

Table 7: Market scoring rubric

Criterion	Description	Weight (w_i)	Rating (1–5) Guidance
Governance & decentralization	Openness, accountability, resilience	20%	1 = centralized/opaque; 5 = credibly neutral
Identity/attribution & interoperability	DID/VC support, ORCID/CRediT, schema alignment	18%	1 = siloed; 5 = standards-based
Data/compute & verifiability	FAIR, registries, verifiable compute/storage	22%	1 = minimal; 5 = verifiable/reproducible
Incentives/tokenomics & sustainability	Funding, reviewer incentives, runway	12%	1 = unclear; 5 = robust/ sustainable
UX/onboarding & adoption	Friction, documentation, integrations	18%	1 = high friction; 5 = smooth/integrated
Compliance/ethics & IP posture	Licensing, GDPR, data sovereignty	10%	1 = risks unmanaged; 5 = robust compliance

Criterion 1: Governance & decentralization

- Centralized/opaque (score=1): Single admin or closed committee; no public roadmap; upgrade keys concentrated; limited transparency.
- Mixed/hybrid (score=3): Documented processes; multisig/DAO with clear proposals but notable centralization or weak voter participation.
- Credibly neutral (score=5): Open governance with broad participation; on-chain proposals, published policies/metrics; minimized trust in maintainers; resilience plans.

Evidence: Governance docs, on-chain records, multisig policies, auditability of processes, decentralization metrics.

Criterion 2: Identity/attribution & interoperability

- Siloed (score=1): No ORCID/CRediT; no DID/VC; proprietary IDs.
- Partial (score=3): ORCID integration or DID/VC; limited schema mapping; some export.
- Standards-based (score=5): ORCID + CRediT + DID/VC; mappable schemas; cross-system verification; well-documented APIs/SDKs.

Evidence: API docs, schema mappings, standards compliance statements, interop demos.

Criterion 3: Data/compute & verifiability

- Minimal (score=1): Ad-hoc storage; no provenance; no verifiability.

- Adequate (score=3): FAIR-leaning practices; basic provenance; reproducibility possible with effort.
- Verifiable/reproducible (score=5): FAIR-aligned registries; tamper-evident provenance; verifiable storage/compute; reproducible pipelines with artifacts.

Evidence: FAIR checklists, provenance registry, compute proofs, replication artifacts, SLAs.

Criterion 4: Incentives/tokenomics & sustainability

- Unclear/fragile (score=1): No coherent incentives; short runway; extractive fee model.
- Reasonable (score=3): Some reviewer/funder incentives; transparent treasury; medium runway; basic anti-gaming.
- Robust/sustainable (score=5): Well-specified incentives (e.g., retro/PGF); empirically validated; diversified funding; explicit sustainability and abuse-resistance.

Evidence: Token/treasury docs, grants records, usage incentives, stress tests, fiscal reporting.

Criterion 5: UX/onboarding & adoption

- High friction (score=1): Poor docs; complex flows; few integrations; low adoption signals.
- Usable (score=3): Clear docs; acceptable flows; some integrations; emerging adoption.
- Smooth/integrated (score=5): Excellent docs; low-friction onboarding; broad integrations; strong adoption/retention evidence (case studies, metrics).

Evidence: Onboarding flows, developer docs, SDKs, observed integrations, usage/retention metrics.

Criterion 6: Compliance/ethics & IP posture

- Risks unmanaged (score=1): No licensing clarity; no GDPR/data-sovereignty stance; ethics not addressed.
- Managed (score=3): Standard OSS licenses; basic GDPR controls; ethics statement; DPIA in progress.
- Robust compliance (score=5): Clear IP/licensing for research artifacts; GDPR/DPIA and data-residency options; responsible research policy and audits.

Evidence: Licensing files, privacy policies/DPIAs, data-processing terms, ethics policies.

13 Appendix D. Detailed comparison tables and provider profiles

Method. Providers are scored on six criteria using one to five anchors and the weights in Appendix C. The criteria are:

- Governance and decentralization (w = 0.20)
- Identity and attribution plus interoperability (w = 0.18)
- Data and compute verifiability (w = 0.22)
- Incentives and sustainability (w = 0.12)
- UX and adoption (w = 0.18)
- Compliance, ethics, and intellectual property (w = 0.10).

The composite score S equals the weighted sum and ranges from one to five. Interpretation bands are as follows, 4-5 Leading, 3-3.9 Advancing, 2-2.9 Emerging, below 2 Nascent.

Table 8: D1. Funding communities and BioDAOs

Provider	C1 Gov	C2 Identity	C3 Verifiability	C4 Incentives	C5 UX	C6 Compliance	S	Category
VitaDAO	4	3	2	4	3	3	3.10	Advancing
HairDAO	3	2	2	3	3	2	2.50	Emerging
ValleyDAO	4	3	2	4	3	3	3.10	Advancing
AthenaDAO	3	2	2	3	3	3	2.60	Emerging
PsyDAO	3	2	1	3	3	2	2.28	Emerging
Bio Protocol (BIO)	3	2	2	3	3	2	2.50	Emerging
LabDAO	3	2	2	3	3	2	2.50	Emerging

Table 9: D1. Adoption signals for D1

Metric	Definition	Source	Collection method	Notes
Grants funded per quarter	Approved grants and total amount by quarter	https://www.vitadao.com/researchers	Manual scrape, program API	Normalized by treasury size
Decision latency	Median days from proposal to decision	https://www.vitadao.com	Manual scrape	Timestamps for proposals and votes
Voter participation	Eligible voters and turnout for program or policy votes	https://www.vitadao.com	Manual scrape	Turnout as share of eligible

Outputs ratio	Share of funded projects producing public artifacts (preprints, datasets, code)	Program pages across providers	Manual review	Link outputs to DOIs where present

Table 10: VitaDAO

Source type	URL	Access date
Program and research pages	https://www.vitadao.com/researchers	2025-10-07
Fellowship or grants overview	https://www.vitadao.com/fellowship	2025-10-07
Community hub	https://www.vitadao.com	2025-10-07

Table 11: HairDAO

Source type	URL	Access date
Project site	https://www.hairdao.xyz	2025-10-07
Community or docs	https://www.hairdao.xyz	2025-10-07

Table 12: ValleyDAO

Source type	URL	Access date
Project site	https://www.valleydao.bio	2025-10-07
Programs or grants	https://www.valleydao.bio	2025-10-07

Table 13: AthenaDAO

Source type	URL	Access date
Project site	https://www.athenadao.co	2025-10-07
Programs or grants	https://www.athenadao.co	2025-10-07

Table 14: PsyDAO

Source type	URL	Access date

Project site	https://www.psydao.io	2025-10-07
Launch or program pages	https://www.psydao.io	2025-10-07

Table 15: Bio Protocol (BIO)

Source type	URL	Access date
Project site	https://www.bioprotocol.io	2025-10-07
Overview	https://www.bioprotocol.io	2025-10-07

Table 16: LabDAO

Source type	URL	Access date
Project site	https://www.labdao.xyz	2025-10-07
Docs or community	https://www.labdao.xyz	2025-10-07

Table 17: D2. Data and knowledge infrastructure

Provider	**C1 Gov**	**C2 Identity**	**C3 Verifiability**	**C4 Incentives**	**C5 UX**	**C6 Compliance**	**S**	**Category**
OriginTrail DKG	4	4	4	3	3	3	3.60	Advancing
Galeon	3	3	4	3	3	4	3.32	Advancing
Open Research Knowledge Graph (ORKG)	4	4	3	2	3	3	3.26	Advancing

Table 18: Adoption signals for D2

Metric	Definition	Source	Collection method	Notes
API coverage and stability	Endpoints, versioning, rate limits, uptime proxies	https://docs.origintrail.io	Manual review and ping	Record version and limits
Documented integrations	Repositories, EHRs or third party systems integrated	https://www.galeon.care	Manual review	Capture partner names
Consent and data use metadata	Presence of consent or data use fields in schemas	https://www.galeon.care	Manual review	Record schema versions

Table 19: OriginTrail DKG

Source type	URL	Access date
Docs	https://docs.origintrail.io	2025-10-07
Technology overview	https://origintrail.io/technology/decentralized-knowledge-graph	2025-10-07

Table 20: Galeon

Source type	URL	Access date
Platform site	https://www.galeon.care	2025-10-07
Concept or whitepaper pages	https://www.galeon.care	2025-10-07

Table 21: Open Research Knowledge Graph (ORKG)

Source type	URL	Access date
Platform site	https://www.orkg.org	2025-10-07
API or docs	https://www.orkg.org	2025-10-07

Table 22: D3. Attribution and integrators

Provider	**C1 Gov**	**C2 Identity**	**C3 Verifiability**	**C4 Incentives**	**C5 UX**	**C6 Compliance**	**S**	**Category**
ResearchHub and ResearchCoin (RSC)	3	3	2	3	3	2	2.68	Emerging
OpenAlex	4	4	3	2	4	3	3.44	Advancing
Crossref	4	4	2	2	4	4	3.32	Advancing
DataCite	4	4	3	2	4	4	3.54	Advancing
OpenReview	4	4	3	2	4	3	3.44	Advancing
PREreview	3	3	2	2	4	3	2.84	Emerging

Table 23: Adoption signals for D3

Metric	Definition	Source	Collection method	Notes

Identifier write backs	Presence of ORCID and DOI write backs or exports	https://docs.openalex.org	Manual review	Scope and limits
API freshness and volume	Update cadence, entity counts, API throughput	https://docs.openalex.org	Manual review	Daily update policy
Public curation volume	Number of public reviews or curations per month	https://docs.researchhub.com	Manual review	Use visible counters

Table 24: ResearchHub and ResearchCoin (RSC)

Source type	URL	Access date
RSC overview	https://docs.researchhub.com/researchcoin/what-is-researchcoin	2025-10-07
Foundation or about	https://researchhub.foundation/research-coin	2025-10-07

Table 25: OpenAlex

Source type	URL	Access date
API overview	https://docs.openalex.org/how-to-use-the-api/api-overview	2025-10-07
Entities and schema	https://docs.openalex.org/about-the-data/what-entities-are-available	2025-10-07

Table 26: Crossref

Source type	URL	Access date
REST API docs	https://www.crossref.org/documentation/retrieve-metadata/rest-api/	2025-10-07
Metadata plus	https://www.crossref.org/documentation/metadata-plus/	2025-10-07

Table 27: DataCite

Source type	URL	Access date
REST API	https://support.datacite.org/docs/api	2025-10-07
Metadata schema	https://schema.datacite.org	2025-10-07

Table 28: OpenReview

Source type	URL	Access date
Platform site	https://openreview.net	2025-10-07
FAQ and docs	https://openreview.net/faq	2025-10-07

Table 29: PREreview

Source type	URL	Access date
Platform site	https://www.prereview.org	2025-10-07
About	https://www.prereview.org/about	2025-10-07

Table 30: D4. Storage and compute substrate

Provider	**C1 Gov**	**C2 Identity**	**C3 Verifiability**	**C4 Incentives**	**C5 UX**	**C6 Compliance**	**S**	**Category**
IPFS	4	3	4	2	3	2	3.20	Advancing
Filecoin	4	3	4	3	3	2	3.32	Advancing
Arweave	3	2	3	3	3	2	2.72	Emerging
Ocean Protocol (compute to data)	4	3	4	3	3	3	3.42	Advancing
Bacalhau (compute over data)	3	3	4	2	3	2	3.00	Advancing
Akash Network (decentralized compute)	3	2	3	2	3	2	2.60	Emerging

Table 31: D4. Adoption signals for D4

Metric	Definition	Source	Collection method	Notes
PID plus fixity coverage	Share of artifacts with PID to CID binding and checksums	https://docs.ipfs.tech	Manual spot checks	Sample by repository
Durability commitments	Presence of paid renewal or endowment policies	https://filecoin.io/filecoin.pdf	Manual review	Note policy details

Execution provenance availability	Signed run records exposed to end users	https://www.bacalhau.org	Manual review	Check job metadata

Table 32: IPFS

Source type	URL	Access date
Whitepaper	https://arxiv.org/abs/1407.3561	2025-10-07
Docs	https://docs.ipfs.tech	2025-10-07

Table 33: Filecoin

Source type	URL	Access date
Whitepaper	https://filecoin.io/filecoin.pdf	2025-10-07
Docs	https://docs.filecoin.io	2025-10-07

Table 34: Arweave

Source type	URL	Access date
Yellow paper	https://www.arweave.org/yellow-paper.pdf	2025-10-07
Docs	https://docs.arweave.org	2025-10-07

Table 35: Ocean Protocol (compute to data)

Source type	URL	Access date
Overview	https://oceanprotocol.com	2025-10-07
Docs	https://docs.oceanprotocol.com	2025-10-07

Table 36: Bacalhau (compute over data)

Source type	URL	Access date
Overview	https://www.bacalhau.org	2025-10-07
Docs	https://docs.bacalhau.org	2025-10-07

Table 37: Akash Network (decentralized compute)

Source type	URL	Access date

Overview	https://akash.network	2025-10-07
Docs	https://docs.akash.network	2025-10-07

14 Appendix E. Glossary and abbreviations

ACM Digital Library (ACM DL)
A bibliographic and full text database for computing research, used in this study as a source for literature searches.

API (Application Programming Interface)
A documented interface that allows software to read or write data programmatically. Stable, versioned APIs are essential for interoperability and evidence collection.

Arweave
A storage protocol that offers an endowment based persistence model for long term availability of content addressed artifacts.

AthenaDAO
A community that funds research in women's health, operating with token enabled governance and public program materials.

BioDAO
A domain focused research funding community that coordinates grants, priorities, and translation using token governed processes.

Compute to data
A pattern where code is executed in the environment where data resides, so that sensitive data do not leave controlled locations. Often paired with auditable access logs.

Consent metadata
Structured information that describes what uses of data are permitted, under which conditions, and with which approvals.

Content addressing
A method of identifying data by a cryptographic digest of its content rather than by location. Supports fixity checks and verifiable retrieval.

CRediT (Contributor Roles Taxonomy)
A taxonomy that specifies contribution roles, for example conceptualization or data curation, used to improve attribution beyond authorship order.

CRIS or RIMS (Current Research Information System or Research Information Management System)
Institutional systems that track research outputs, projects, and affiliations, often integrated with PIDs and repository workflows.

Crossref
A DOI registration agency and metadata service for scholarly publications with a public REST API.

DAO (Decentralized Autonomous Organization)
A governance arrangement encoded in software, with rules for proposals, voting, and treasury management, used by several research funding communities.

DataCite
A DOI registration agency for datasets and other research objects with modern API endpoints and schema guidance.

Data use conditions
Machine readable statements that describe how data may be accessed or reused, for example by license, jurisdiction, or consent.

DID (Decentralized Identifier)
A globally unique identifier that can be resolved without a central registry. Often used with verifiable credentials to express portable claims.

DPIA (Data Protection Impact Assessment)
A structured assessment of privacy risks and mitigations for processing activities, required for certain data flows under GDPR.

FAIR principles
Guidelines for making research objects findable, accessible, interoperable, and reusable, with emphasis on machine actionability.

Filecoin
A decentralized storage network that provides economic commitments for data durability through proofs of replication and spacetime.

Fixity
A property of data integrity established by checksums or cryptographic digests so that accidental or malicious changes can be detected.

Galeon
A healthcare data platform that integrates electronic health records and explores privacy preserving model training and compute to data concepts.

GDPR (General Data Protection Regulation)
The European Union regulation that governs personal data protection, including principles, data subject rights, and obligations for controllers and processors.

Governance
The structures and processes that allocate decision rights, define roles, and record changes and appeals. In this report it includes rules for proposals, voter participation, and treasury use.

HairDAO
A patient centered research community that funds early stage work on hair loss and related treatments.

Hypercerts
A data model for representing impact claims that can be recognized by multiple funders without double counting.

Identity
The set of identifiers and claims that link a researcher, an institution, or a system to actions and artifacts across platforms.

IEEE Xplore
A digital library for electrical engineering and computer science literature, used as a source in the search protocol.

Interoperability
The capacity of systems to exchange and use information consistently through standards, stable APIs, and shared schemas.

IPFS (InterPlanetary File System)
A protocol for content addressed transport and verifiable retrieval that underpins decentralized storage ecosystems.

Multisig
A mechanism where multiple keys are required to authorize a transaction. Often used for shared treasury control or upgrade keys.

OpenAlex
An open scholarly knowledge graph with a free API that indexes works, authors, venues, institutions, and concepts.

OpenAPI
A machine readable specification format for describing REST APIs, used to document endpoints, inputs, and outputs.

Open Science Framework (OSF)
A platform for project management, preregistration, and sharing of research artifacts, used here as a source for grey literature.

ORCID
A persistent identifier for researchers, with OAuth and record APIs that support attribution and linking of contributions.

OriginTrail, Decentralized Knowledge Graph (DKG)
An open network for publishing verifiable knowledge assets that can be resolved across multiple chains.

PID (Persistent Identifier)
A stable identifier for an entity, such as a DOI for a dataset or publication, or an ORCID for a person.

PsyDAO
A community that funds psychedelic science using token governed processes and public program materials.

PRISMA ScR
A reporting guideline for scoping reviews that structures identification, screening, eligibility, and inclusion, together with flow counts.

PROV and PROV O
A conceptual model and ontology from W3C for representing provenance as entities, activities, and agents, with typed relations among them.

Provenance
A machine readable account of how a result was produced, including inputs, processes, agents, and outputs, so that claims can be traced and verified.

Quadratic funding
A mechanism for allocating matching funds that amplifies broad community support for public goods.

Replication
The independent repeating of an analysis or experiment using the original data and methods to check consistency of results.

Reproducibility
The ability to obtain the same results when running the same code on the same data, under specified conditions and environment settings.

ResearchCoin (RSC)
A token associated with ResearchHub that is used to reward contribution and curation activities.

ResearchHub
A platform for discussion, curation, and incentives around research outputs, which uses ResearchCoin to reward activity.

Research object
A package that binds data, code, configuration, environment, and metadata into a coherent and citable unit.

RetroPGF (Retroactive Public Goods Funding)
An allocation model that rewards projects after impact has been demonstrated.

RIMS
See CRIS. The terms CRIS and RIMS are used interchangeably in this report.

RO Crate
A community profile for packaging research objects on the web, with human and machine readable manifests.

Runway
The time period that a treasury or project can operate given current reserves and spending plans.

Scopus
A bibliographic database used for the literature search, with coverage across disciplines.

SSRN
A repository for preprints and working papers, used here as a source for grey literature.

Tee, Trusted Execution Environment (TEE)
A hardware backed secure execution context that can provide attestations about code and environment measurements.

Treasure, Treasury
The pool of funds controlled by a community or organization, subject to governance rules and transparency policies.

Verifiable compute
A property of computation where there is evidence that a job was executed with the specified code and inputs, in a specified environment.

Verifiable Credentials (VCs)
Cryptographically signed, portable claims about an entity, for example a reviewer contribution or a dataset stewardship role, that can be selectively disclosed and revoked.

VitaDAO
A community that funds longevity science through open calls and portfolio processes, with public governance and program materials.

Web of Science
A multidisciplinary bibliographic index used in the literature search protocol.

W3C (World Wide Web Consortium)
The standards body responsible for many web specifications, including DID, VC, and PROV.

Zenodo
A general purpose repository operated by CERN that issues DOIs for deposited artifacts, referenced in comparative discussions of repository practices.

The Project is financed by the Republic of Cyprus through the Research and Innovation Foundation