

The Federated Data Space Infrastructure of Materials Commons: Concept and Architecture

Simon Stier^{1,*}, Marek Cebecauer², Jakub Klímek^{2,3}, Jörg Schaarschmidt⁵, Andreas Räder¹, Natalia Bedoya-Martinez⁴, Rickard Armiento⁶ and Lukas Gold¹

¹Fraunhofer Institute for Silicate Research ISC, Neunerplatz 2, 97082 Würzburg, Germany

²J. Heyrovský Institute of Physical Chemistry of the Czech Academy of Sciences, Prague, Czech Republic

³Charles University, Faculty of Mathematics and Physics, Prague, Czech Republic

⁴Materials Center Leoben Forschung GmbH, Leoben, Austria

⁵Karlsruhe Institute of Technology, Karlsruhe, Germany

⁶Linköping University, Department of Physics, Chemistry and Biology (IFM), Linköping, Sweden

Abstract

Materials research and development produce data and services that remain locked in institutional silos: repositories, electronic lab notebooks, simulation codes, and increasingly autonomous experiments that are individually valuable but mutually invisible. Connecting these heterogeneous resources requires a common framework through which they can be discovered, accessed, and used across institutional boundaries without giving up local control. Industrial data space initiatives have converged on the Data Space Protocol (DSP) as a standard for sovereign data exchange, yet their reference implementations presume dedicated IT operations and governance structures that scientific labs do not have, while simultaneously focusing on the exchange of well-defined static assets. The European MaterialsCommons initiative has been established to bridge this gap while accommodating the heterogeneous resources and capabilities of scientific participants. We present the technical perspective of the initiative that aims to establish a federated data space built on (i) a layered participation model in which the cost of joining the data space scales with the capability of the participant, from publishing a single static catalog document to operating a full industrial connector, (ii) domain extensions to the catalog layer, namely MaterialsDCAT-AP, a domain-specific application profile of DCAT-AP, and a service catalog that describes dynamic assets such as database queries, computational workflows, and remote experiments, and (iii) a replicable node architecture with federated identity that lets scientists participate with their home-institution accounts and existing tools. We illustrate the architecture through three scientist journeys and discuss open challenges.

Keywords

data spaces, DCAT-AP, federated infrastructure, materials science, FAIR data, knowledge graphs, semantic interoperability

1. Introduction

The European MaterialsCommons initiative aims to let any materials scientist discover, access, and contribute data and services across institutional boundaries as easily as browsing the web. However, the current reality is far from it. Measurement data sits in instrument-specific files on lab shares, curated datasets in disconnected repositories, simulation workflows on institutional clusters, and emerging self-driving labs behind bespoke interfaces. Each connection between two of these systems is negotiated and engineered individually.

Industry faces the same problem and has responded with data spaces: federations of sovereign participants that exchange data under common protocols and policies rather than through a central platform. The Data Space Protocol (DSP) now provides an interoperable, standardized baseline [1], embedded in the architecture and governance frameworks of IDSA and Gaia-X [2, 3]. However, the

SeMatS 2026: The 3rd International Workshop on Semantic Materials Science co-located with the 25th International Semantic Web Conference (ISWC 2026), October 25th, 2026, Bari, Italy

*Corresponding author.

✉ simon.stier@isc.fraunhofer.de (S. Stier); rickard.armiento@liu.se (R. Armiento)

ORCID: 0000-0003-0410-3616 (S. Stier); 0000-0002-4606-1218 (M. Cebecauer); 0000-0001-7234-3051 (J. Klímek);

0000-0002-4389-2366 (J. Schaarschmidt); 0000-0002-2673-0832 (A. Räder); 0000-0001-9824-6129 (N. Bedoya-Martinez);

0000-0002-5571-0814 (R. Armiento); 0000-0001-7444-2969 (L. Gold)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

resulting reference implementations target enterprises with dedicated IT departments. Two mismatches follow. The first is organizational: a university lab or research group is usually not a legal entity in its own right, so the contractual and governance procedures that industrial data spaces assume of a participant have no counterpart on the academic side. The second is operational: the mismatch concerns who has to run a service rather than any particular technology. Joining an industrial data space today means deploying and continuously operating a server-side component that must stay reachable, patched and authenticated, which presupposes staff whose job it is. A research group has no such staff; what it does have is the equipment of ordinary scientific work, namely storage it can write files to, an interpreter it can run a client library in, a browser, and possibly an Electronic Lab Notebook (ELN). An entry level that requires operating a service therefore excludes exactly the participants the federation most needs, whereas one that requires only publishing a file, or invoking a library, does not. Moreover, scientific assets are not only static files: the most valuable offers are dynamic, including live databases, computational workflows on HPC systems, and experimental services up to fully autonomous laboratories.

We describe how the European MaterialsCommons initiative projects the industrial data space blueprint onto an infrastructure that works for both scientists and industry. Our contributions are:

1. a layered participation model that decouples joining the federation from operating connector infrastructure (Section 3);
2. domain extensions to the catalog layer: a MaterialsDCAT-AP and a service catalog for dynamic assets (Section 4);
3. a replicable node architecture with federated identity, packaged for adoption in whole or in parts (Section 5).

This paper addresses the exchange and discovery layer, i.e. how assets become visible and accessible across institutions. How an individual dataset and its payload are described semantically is treated in separate work on Object-Oriented Linked Data (OO-LD) profiles [4]; the concern here is how the federation accommodates differences in the descriptions exposed by individual providers, which Section 4.2 addresses at the discovery layer.

2. Background

Data space frameworks. The International Data Spaces Association (IDSA) and Gaia-X define architectures and governance for sovereign data exchange [2, 3]. The Data Space Protocol standardizes the interaction sequence between participants: catalog exchange (what is offered), contract negotiation (under which policy), and the transfer process (how access to the bytes is coordinated) [1]. Eclipse Dataspace Components (EDC) is the de facto reference implementation, deployed at scale in industrial federations such as Catena-X [5]. The Data Spaces Support Centre consolidates these building blocks in its Data Spaces Blueprint [6].

Research data infrastructure. On the science side, the European Open Science Cloud federates generic research data services [7]. In materials science, repositories and platforms such as NOMAD [8], Materials Cloud [9], and the MaterialDigital data portal [10] are established; research initiatives such as CAPeX [11] are building dedicated data infrastructures; and the OPTIMADE consortium has demonstrated that a common query API across dozens of structure databases is achievable [12, 13]. At the catalog level, the MaterialDigital portal already builds on a CKAN-based software stack that federates dataset descriptions using DCAT, the W3C standard vocabulary for data catalogs, and the initiative has consolidated its concepts for a semantically accessible materials data space [14]. The European application profile of DCAT, DCAT-AP, is extended by domain profiles such as GeoDCAT-AP for geospatial and MLDCAT-AP for machine learning assets [15].

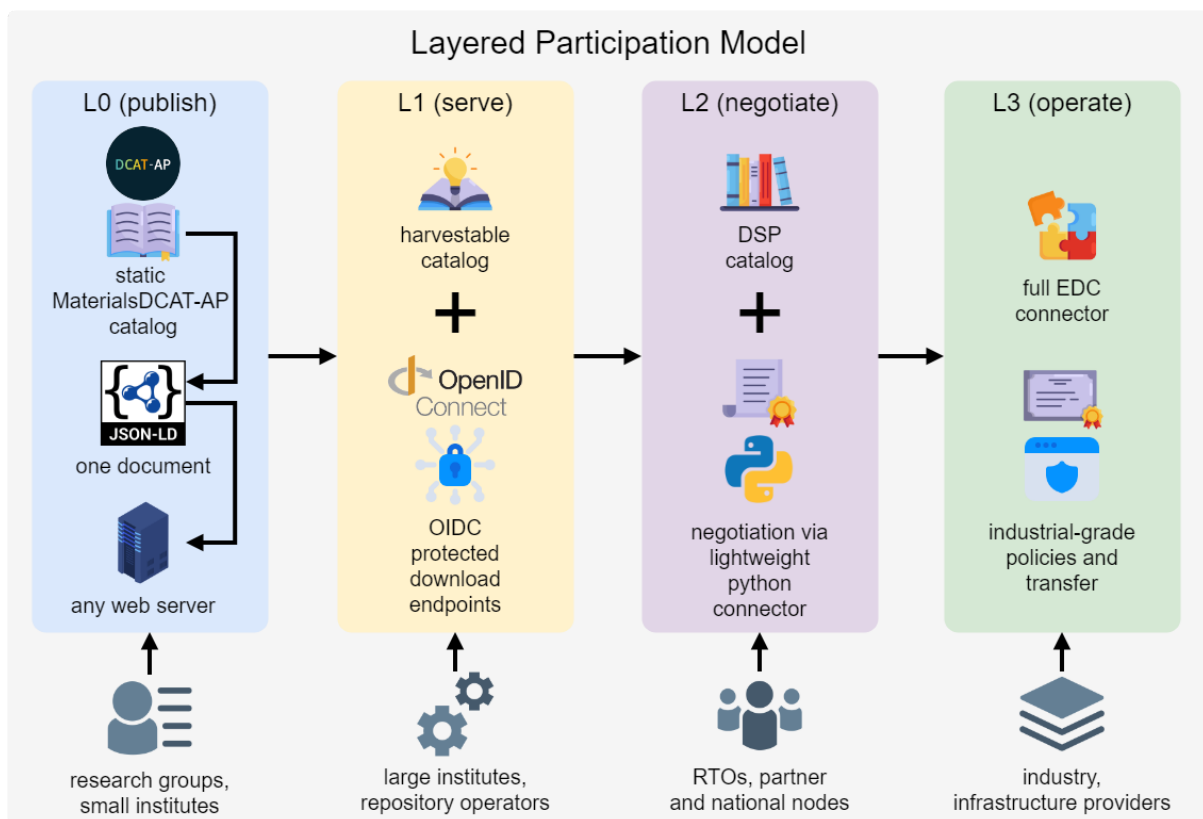


Figure 1: Layered participation model. Capability and operational cost grow from left to right; the bottom row names the typical entry stakeholders per level, each free to migrate upward as its sharing scenarios demand; assets from every level end up in the same federated index, whether harvested or federated from the provider.

The gap. Overarching architectures exist in both data-space frameworks and research data infrastructures, but their adoption in domain research infrastructures is uneven, and ready-to-use reference implementations that a research group can realistically operate are scarce. Ontology-based data spaces have been demonstrated within single projects, for example for battery production [16] and for steel and copper development [17], but remain scoped to a fixed consortium and project lifetime rather than an openly joinable federation. Bridging this gap is a design problem, not merely an integration effort: the participation cost, the catalog expressiveness, and the deployment model all need to be rethought for scientific practice.

3. Layered Participation Model

The central design decision is that the entry barrier must scale with the capability of the participant, not with the ambition of the data space. We therefore define four participation levels against the three DSP pillars (catalog, contract, transfer), shown in Figure 1. A participant enters at the lowest level that matches its resources and migrates upward only when its sharing scenarios demand it.

L0 (publish). A lab describes its assets in a single static MaterialsDCAT-AP catalog document, a JSON-LD file published as part of the dataset or as a reference to existing public datasets, served from any web server, repository landing page, or even a Git repository. No running service is required, but catalog and assets should be resolvable at stable public locations; hosting on GitHub, Zenodo, or a discipline-specific repository, the typical academic case, suffices. The federation’s harvesters pick up the document and make the assets discoverable, following the publish-then-harvest pattern of the FAIR Data Point [18], but lowering the entry barrier by requiring only a single static file rather than a

running metadata service. We defined this level because the first hurdle, being visible at all, must cost close to nothing.

L1 (serve). The participant additionally exposes download or query endpoints, protected where access is restricted by standard OpenID Connect (OIDC) tokens, while the licence under which an asset may be used is stated in the catalog entry rather than enforced by the endpoint. Because the participant now controls a server, one convention lets the federation find the catalog without being configured for each provider: a small discovery document sits at a path fixed in advance, under the `/well-known/` prefix defined by RFC 8615 [26], and names the absolute location of the catalog, which clients follow rather than construct. Off-the-shelf web servers and repository software suffice; no data-space-specific component is involved yet.

L2 (negotiate). The participant speaks DSP itself: its catalog is served via the protocol, and access is granted through contract negotiation. What changes at this level is how the terms are represented. Below it, a licence is represented by a link to a human-readable document, typically a Creative Commons deed; the protocol instead expects a policy expressed in ODRL [25], an RDF vocabulary a machine can evaluate. Open assets exploit this: their policy grants access unconditionally, so agreement is automatic and no human step is involved, while the exchange still takes place inside the protocol rather than beside it. To make this level reachable for scientists, we are building a lightweight, pip-installable Python connector (“EDC-Light”) that implements the required protocol subset and fits the Python-centric tooling of the community.

L3 (operate). Industrial partners and infrastructure providers operate full EDC deployments with fine-grained usage policies, auditability, and managed transfer channels.

The participation level is a property of the participant, not of the data space. One federation therefore spans all four levels, and a consumer never has to know where a provider sits. The two connector tracks differ in what they demand of the participant rather than in which language they are written: the reference implementation is a service to be deployed and operated, while the lightweight one is a library imported into an existing analysis environment. Keeping the two verifiably compatible is part of the infrastructure work. In a DSP catalog every distribution is reached through a data service, so from L2 upward a connector must exist for a provider’s assets, but nothing requires the provider to be the one running it. Small and medium-sized enterprises that cannot afford self-operation delegate connector operation to connector-as-a-service providers, a pattern established in Catena-X by its operating company Cofinity-X and certified enablement service providers; the same route keeps the step up from a published catalog small for a research group. Across all levels the assets and the properties describing them stay the same, while the serialization follows whichever profile the transport imposes, so conformance can be stated as a checklist a participant runs against its own endpoint [24].

4. Domain Extensions to the Catalog Layer

The catalog is where a generic data space stops being useful for materials science unless it is extended. Generic DCAT-AP describes datasets and services, but not the materials- and technology-specific contexts.

4.1. MaterialsDCAT-AP

Catalog entries must answer the questions materials scientists ask: which material system, what kind of structure, which synthesis or processing route, which characterization or simulation method, under which conditions? The implementation and establishment of an official MaterialsDCAT-AP is a committed deliverable of the MaterialsCommons project. This MaterialsDCAT-AP will extend DCAT-AP with these descriptors and anchor them in the domain ontology ecosystem, in particular EMMO [19]

and PMDco [20], so that catalog entries are not free-text but machine-interpretable and validatable. Validation rules make conformance testable at harvest time; entries that do not validate are flagged rather than silently degraded. The profile follows the methodological precedent of GeoDCAT-AP: extend, do not fork, so that every MaterialsDCAT-AP record remains a valid DCAT-AP record and generic European data portals can consume it; ChemDCAT-AP applies a similar extension pattern to chemistry and catalysis [21]. Indicative coverage spans nanomaterials, metals and polymer composites, as well as materials defined by their application, such as photovoltaic and battery materials; the profile is intended for submission to an open standardization body.

Concretely, a battery cycling dataset published this way carries the cell chemistry and format as a material system, the cell assembly route as its processing history, the electrochemical cycling protocol as its characterization method, and the temperature and rate window as conditions, each as an IRI drawn from the domain ontologies rather than as a free-text field. That is what makes the difference between a catalog entry a search engine can index and one a materials scientist can query for cells of a given chemistry cycled under a given protocol.

4.2. Semantic Interoperability Across the Federation

Describing assets uniformly is not the same as making their contents comparable, and the two problems are solved at different layers. The catalog layer, above, standardizes how an asset is announced. It does not standardize the data inside the asset, which each provider continues to model with the vocabulary of its own community, so a federated query that ranges over several providers meets EMMO, PMDco and various in-house terms for what a materials scientist regards as the same quantity.

We do not mandate one ontology. The catalog profile itself is anchored: MaterialsDCAT-AP descriptors resolve to EMMO and PMDco terms, so discovery is semantically uniform even when payloads are not. Payload semantics come from the schema that a distribution declares conformance to, following the companion profile work [4], so a consumer that retrieves a dataset also retrieves the definitions it needs to read it, down to the unit of each measured value. Where two communities describe the same property with different terms, the reconciliation is published as a mapping set instead of being resolved by re-modeling the data, which keeps the provider's own representation intact and lets the consumer choose the reading it needs. Anchored descriptors, declared payload schemas and published mapping sets are all indexed by the federated knowledge graph, so that the semantic gap between providers is made explicit and queryable instead of being hidden behind a lowest common denominator.

4.3. Service Catalog for Dynamic Assets

Scientific assets are services, not only files. A live database answering structure queries, a simulation workflow executable on HPC resources, an analysis pipeline, or a remote experiment up to a self-driving laboratory cannot be represented as a downloadable distribution. We therefore extend the catalog with service descriptions based on DCAT-AP's data service concept (`dc:DataService`, in DCAT-AP since version 2 following W3C DCAT 2 [15]), following the trail of MLDCAT-AP: every provider answers with a semantic self-description of its offers, ranging from a plain HTTP download to SPARQL, GraphQL, or REST endpoints, including the schemas of inputs and outputs. Self-descriptions are aggregated in thematic API gateways: a gateway bundles the related offers of one theme, for example battery databases or simulation services, behind a single endpoint that forwards queries to the registered providers and merges the results, so consumers address one endpoint per theme rather than every provider individually. OPTIMADE [12] is the proof of concept for this pattern in materials science, federating structure queries across dozens of databases through one theme-level API [13]; the service catalog generalizes it from a single hand-standardized interface to self-described services of any kind, discovered through the catalog rather than fixed by specification.

The knowledge graph introduced above is the discovery backbone across all participation levels, indexing datasets, services, tools, vocabularies and the organizations behind them, and searchable by semantic query, graphical navigation and natural language. It covers fully public assets and assets

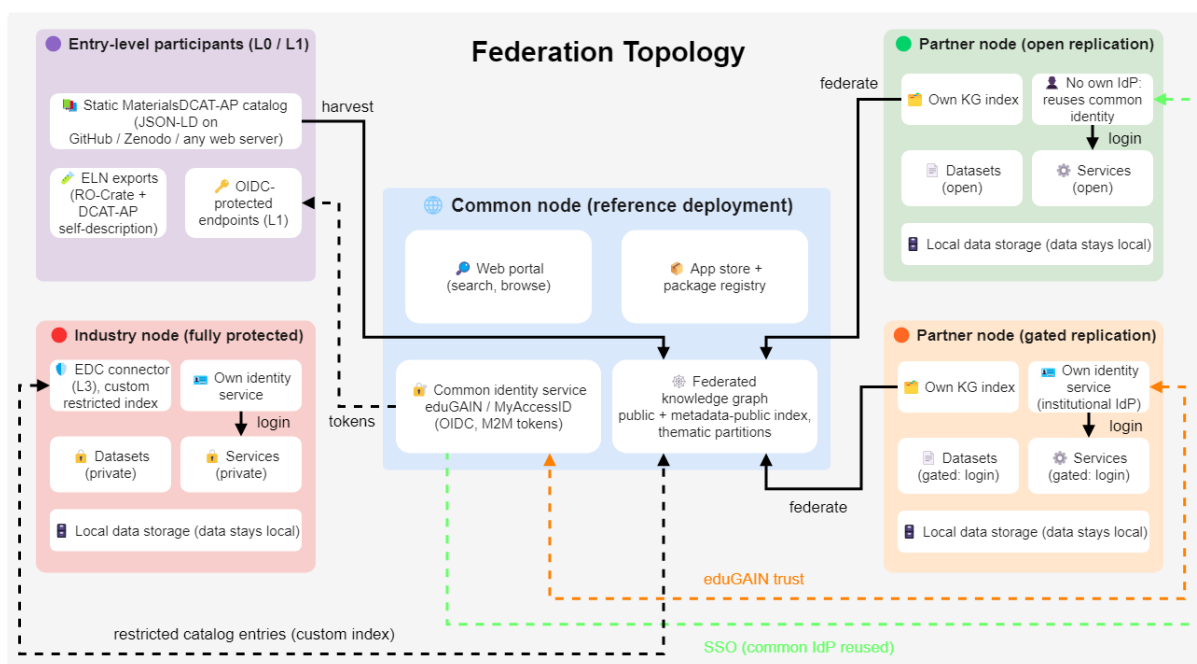


Figure 2: Federation topology with three replication variants. A common reference node provides portal, knowledge graph, app store, and a common identity service. An open partner node replicates the stack, exposes open datasets and services, and reuses the common identity; a gated partner node runs its own institutional identity provider, joined through the eduGAIN trust federation, and requires login for datasets and services; a fully protected industry node keeps assets private behind its own identity and a full EDC connector, contributing at most restricted catalog entries via a custom index. Entry-level participants (L0/L1) are harvested. Data always remains local; only metadata is exchanged.

whose metadata is public; gated catalogs require custom index generation on the provider's side. Detail varies with breadth, so the top-level index stays shallow while thematic partitions carry the domain detail.

The service catalog is what turns the data space into a research platform: discovery of a dataset, of the workflow that can process it, and of the lab that can produce a follow-up sample are all the same kind of query.

5. Replicable Node Architecture for Scientists

The third contribution addresses deployment. A federation only grows if joining it as a node, not just as a participant, is a supported, repeatable operation. Figure 2 shows the resulting topology.

Reproducible stack. The entire node is deployed declaratively on Kubernetes via maintained charts and manifests, complemented by an app-store frontend and a registry of validated container images. For contributors, code-to-container templates wrap existing scripts in predefined web API and authentication layers, with continuous integration for documentation, dependency and license checks, and tests. Interactive and legacy tooling is served, e.g., through Jupyter and remote-desktop containers that can mount remote resources. The aim is that a scientist can turn an existing script into a cataloged, authenticated service without becoming a platform engineer.

Topology and sovereignty. A common reference node operates the shared entry point: portal, knowledge graph, app store, identity. Institutions replicate the same stack as partner or national nodes and are expected to run their own knowledge graph index, which the common node *federates* (query federation and index exchange) rather than harvests; *harvesting* is reserved for entry-level participants

with no infrastructure of their own. Participants must publish metadata, not data: research data may stay on the lab's or institution's own storage; only the catalog metadata must reach the federated knowledge graph or one of the partner catalogs. Where data is replicated for resilience, it is replicated across European core providers, extending the single-center preservation model established by Materials Cloud.

Federated identity. Scientists authenticate with their existing home-institution accounts through eduGAIN and MyAccessID (the GEANT identity and access services), so that no new credentials are created for the federation. Machines, workflows, and agents use token-based machine-to-machine authentication with commonly specified OIDC claims; since DSP does not fix what a token must assert, agreeing these claims across the federation fills a gap the protocol leaves open by design. Access to assets is graded (open, restricted, private), and the same identities carry through from the portal down to HPC access. Every additional account and every manual approval step costs adoption.

Three scientist journeys. The following make the architecture concrete:

- **J1, publish:** an electrochemist finishes a cycling series on a pouch-cell batch and exports it from the ELN as an RO-Crate [22] carrying an embedded MaterialsDCAT-AP self-description, with the cell chemistry, the assembly route and the cycling protocol as ontology IRIs; the file lands on the group's web space (L0) and is harvested into the knowledge graph the same day, at the cost of one export and one upload.
- **J2, consume:** a modeler building a degradation model searches for cells of that chemistry cycled in a given temperature window, finds the dataset in the portal and pulls it into a Python session through the standard tooling; open data requires no account, and no manual normalization either, because the schema already fixes what each column holds and in which unit [4].
- **J3, offer a service:** the modeler wraps the trained capacity-fade predictor with the stack template; continuous integration builds a container, the service self-describes into the catalog with the cell types it is valid for, and it appears in the app store. Acceptable use is enforced at this point: consumers invoke the service after logging in with their academic account.

6. Realization and Outlook

The architecture described here is the technical stack proposed for the reference implementation being realized within the MaterialsCommons initiative, a European innovation action of 26 academic partners and 21 associated partners coordinated by Fraunhofer (2026-2030, Horizon Europe grant 101293545) [23]. The initiative targets a production-grade federated prototype with multiple replicated nodes and integrations with existing infrastructures, including NOMAD/FAIRmat, Materials Cloud, MaterialDigital, and CAPEX [11], behind a single web entry point.

The requirements of Section 3 are written up as a provider-side specification of the entry levels, published as a prototype draft for discussion [24] and derived by checking the model against the Dataspace Protocol, ODRL and DCAT 3. It expresses the levels as four conformance tiers: Tier 0 for a static catalog served over plain HTTP, and Tiers 1 to 3 for a growing subset of the protocol as an early draft for discussion.

What the tiers do not decide is why a provider would join, how an existing industrial stack connects, and who runs the federation once the funding ends.

The first participation level has to pay for itself before it obliges anyone. Our working answer is that L0 providers keep the features that cost the federation nothing to grant and that a lab actually wants: a citable landing page per dataset, inclusion in portal search and in the knowledge graph, and usage statistics for their own records. Gated features begin where the federation carries real cost, which is brokered access and machine-to-machine delivery. Whether that level belongs inside the protocol is still an open discussion: one reading is that a data space begins where an access decision is made,

whereas ours is that it begins at common retrieval, since DCAT standardizes the vocabulary of a catalog without standardizing how that catalog is fetched.

Semantic friction between the industrial stack and the W3C linked data stack has to be bridged rather than argued away [27]. In the Asset Administration Shell, an asset is described by submodels instantiated from templates, which play the role that a catalog profile plays here; the bridge is therefore a published mapping between those templates and the catalog terms, maintained with the same versioning discipline as the profile itself.

Long-term governance and sustainability, that is who operates the common node and curates the catalogs after the funded phase, is a design constraint from day one. Architecturally no single node is load-bearing, since every function the common node performs is either replicable by a partner node or reducible to harvesting a static file, and the intended route is that national nodes are carried by their host institutions as research infrastructure while catalog curation stays with the providers, whose metadata obligation it is anyway. Replaceable infrastructure does not settle custodianship of the shared specification, though. Which body maintains MaterialsDCAT-AP and the tier profiles after the funded phase is open, and it is the more consequential question, since a federation can survive losing an operator but not losing the profile its members conform to.

Neither the participation levels nor the MaterialsDCAT-AP profile can be settled inside a single project, since both have to work for providers who had no part in designing them. That is why the profile is headed for open standardization and why the tier specification is written as testable requirements rather than as a description of our own deployment.

Conclusion. A federated data space becomes usable for science once the entry cost matches what a participant can actually run, the catalog speaks the domain language and describes services as well as files, and operating a node is a repeatable exercise with federated identity already in place. The standards and the software building blocks largely exist; what has been missing is a path into them that a research group can take. This paper proposes one, in the form of a tiered participation model and a stack shaped by how scientific groups work. Whether it holds will be decided by the implementation now under way and by the governance questions above.

Acknowledgements

The authors thank all partners of the MaterialsCommons consortium for the discussions that shaped this work. This work is part of the MaterialsCommons project. Funded by the European Union (Horizon Europe grant agreement No 101293545). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them.

Declaration on Generative AI

During the preparation of this work, the authors used Claude (Anthropic) in order to: structure content, improve writing style and citation management. All concepts, designs, and technical results presented originate from the authors. When using these tools, the authors always validated the results and take full responsibility for the publication's content.

References

- [1] Eclipse Dataspace Working Group: Eclipse Dataspace Protocol. <https://eclipse-dataspace-protocol-base.github.io/DataspaceProtocol/>; under ISO standardization as ISO/IEC DIS 26450. <https://www.iso.org/standard/93502.html>

- [2] International Data Spaces Association: IDS Reference Architecture Model 4 (IDS-RAM 4.2). https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0; <https://internationaldataspaces.org/use/reference-architecture/>
- [3] Gaia-X European Association for Data and Cloud AISBL: Gaia-X Architecture Document, Release 3.1 (2026). <https://docs.gaia-x.eu/technical-committee/architecture-document/latest/>
- [4] Stier, S., et al.: Towards Object-Oriented Linked Data Profiles for Material Science. In: Proceedings of the 3rd International Workshop on Semantic Materials Science (SeMatS 2026), CEUR Workshop Proceedings (2026)
- [5] Eclipse Dataspace Components. <https://github.com/eclipse-edc/Connector>; Catena-X Automotive Network e.V. <https://catena-x.net/>
- [6] Data Spaces Support Centre: Data Spaces Blueprint, version 3.0 (2025). <https://blueprint.dssc.eu/>
- [7] European Open Science Cloud (EOSC). <https://eosc.eu/>
- [8] Draxl, C., Scheffler, M.: The NOMAD laboratory: from data sharing to artificial intelligence. *J. Phys. Mater.* 2, 036001 (2019). <https://doi.org/10.1088/2515-7639/ab13bb>
- [9] Talirz, L., et al.: Materials Cloud, a platform for open computational science. *Sci. Data* 7, 299 (2020). <https://doi.org/10.1038/s41597-020-00637-5>
- [10] Platform MaterialDigital. <https://materialdigital.de/>; PMD data portal tutorial. https://github.com/materialdigital/dataportal_tutorial
- [11] CAPeX, Pioneer Center for Accelerating P2X Materials Discovery. <https://capex.dtu.dk/>
- [12] Andersen, C.W., et al.: OPTIMADE, an API for exchanging materials data. *Sci. Data* 8, 217 (2021). <https://doi.org/10.1038/s41597-021-00974-z>
- [13] Evans, M.L., et al.: Developments and applications of the OPTIMADE API for materials discovery, design, and data exchange. *Digital Discovery* 3(8), 1509-1533 (2024). <https://doi.org/10.1039/d4dd00039k>
- [14] Bayerlein, B., Waitelonis, J., et al.: Concepts for a semantically accessible materials data space: Overview over specific implementations in materials science. *Advanced Engineering Materials* 27(8), 2401092 (2025). <https://doi.org/10.1002/adem.202401092>
- [15] SEMIC: DCAT-AP 3.0.1. <https://semiceu.github.io/DCAT-AP/releases/3.0.1/>; GeoDCAT-AP 3.1.0. <https://semiceu.github.io/GeoDCAT-AP/releases/3.1.0/>; MLDCAT-AP 3.1.0. <https://semiceu.github.io/MLDCAT-AP/releases/3.1.0/>. Data services were introduced in DCAT-AP 2.0.0, aligning with W3C DCAT 2. <https://semiceu.github.io/DCAT-AP/releases/2.0.0/>
- [16] Stier, S.P., Xu, X., Gold, L., Möckel, M.: Ontology-based battery production dataspace and its interweaving with artificial intelligence-empowered data analytics. *Energy Technology* 12(4), 2301305 (2024). <https://doi.org/10.1002/ente.202301305>
- [17] Nahshon, Y., Morand, L., Büschelberger, M., Helm, D., Kumaraswamy, K., Zierrep, P., Weber, M., de Andrés, P.: Semantic Orchestration and Exploitation of Material Data: A Dataspace Solution Demonstrated on Steel and Copper Applications. *Advanced Engineering Materials* 27, 2401448 (2025). <https://doi.org/10.1002/adem.202401448>
- [18] Bonino da Silva Santos, L.O., Burger, K., Kaliyaperumal, R., Wilkinson, M.D.: FAIR Data Point: A FAIR-oriented approach for metadata publication. *Data Intelligence* 5(1), 163-183 (2023). https://doi.org/10.1162/dint_a_00160
- [19] Elementary Multiperspective Material Ontology (EMMO), version 1.0.3. <https://w3id.org/emmo>; <https://github.com/emmo-repo/EMMO>
- [20] Bayerlein, B., et al.: PMD Core Ontology: Achieving semantic interoperability in materials science. *Materials and Design* 237, 112603 (2024). <https://doi.org/10.1016/j.matdes.2023.112603>; <https://w3id.org/pmd/co>
- [21] Strömert, P., et al.: ChemDCAT-AP: Enabling semantic interoperability with a contextual extension of DCAT-AP. *arXiv:2602.01822* (2026), accepted at MTSR 2025.
- [22] Soiland-Reyes, S., et al.: Packaging research artefacts with RO-Crate. *Data Science* 5(2), 97-138 (2022). <https://doi.org/10.3233/DS-210053>
- [23] MaterialsCommons: Federated Digital Infrastructure for European Advanced Materials Research and Development. Horizon Europe grant 101293545 (2026-2030). <https://cordis.europa.eu/project/>

id/101293545

- [24] MaterialsCommons: Materials Commons DSP Tier 1 (Public Datasets and Services) and Materials Commons DCAT-AP GET Protocol Tier 0 (Public Catalogues). Provider-side profiles of the Dataspace Protocol, DCAT-3 and DCAT-AP for open materials data, prototype draft for discussion, version 0.1.0 (2026). <https://materialscommons-staging.github.io/mc-dsp-tiers/0.1.0/>; <https://github.com/MaterialsCommons-Staging/mc-dsp-tiers>
- [25] W3C: ODRL Information Model 2.2. W3C Recommendation (2018). <https://www.w3.org/TR/odrl-model/>
- [26] Nottingham, M.: Well-Known Uniform Resource Identifiers (URIs). RFC 8615, IETF (2019). <https://www.rfc-editor.org/rfc/rfc8615>
- [27] Bader, S.R., Maleshkova, M.: The Semantic Asset Administration Shell. SEMANTiCS 2019, LNCS 11702, 159-174 (2019). https://doi.org/10.1007/978-3-030-33220-4_12