A DCAT 3 Catalog Model for Semantic Interoperability in EOSC

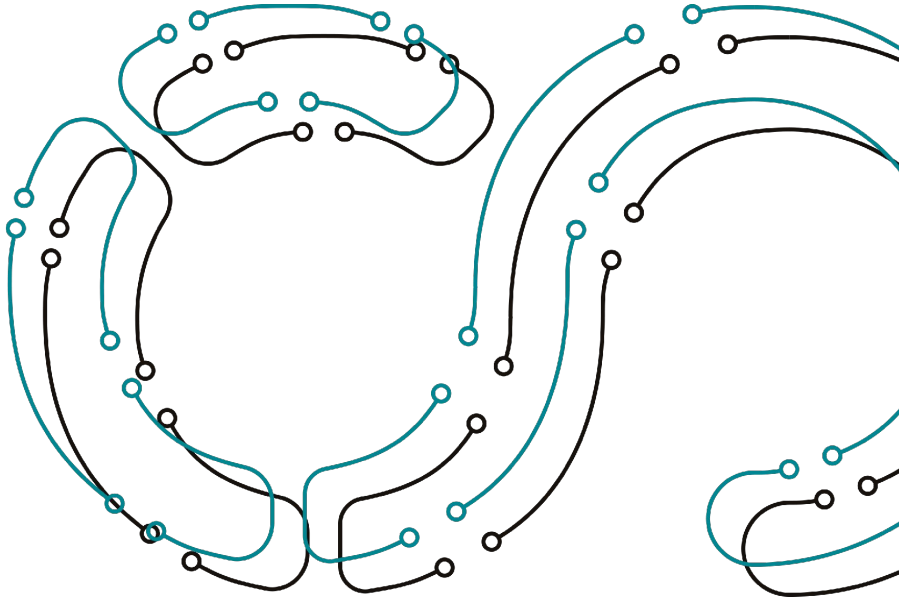
Sharing heterogeneous resources across the Federation without losing information

EOSC Symposium 2026 · Florence · Track 3 — EOSC Federation IF, architecture and data-related policies

Christelle Pierkot · CNRS — Data Terra EOSC Node, deputy coordinator

Julien Homo · Foxcub — Data Terra EOSC Resource Catalogue

Alessandro Rizzo · IRD — Data Terra EOSC Node, coordinator



eosc A multidisciplinary thematic node

Context · Earth & environmental sciences, research products and services

TWO RESEARCH INFRASTRUCTURES

DATA TERRA — French research infrastructure for Earth system data

\+

NFDI4Earth — German NFDI consortium for Earth system sciences

DISCIPLINES

Atmosphere | Ocean | Solid Earth | Land surfaces | Biodiversity

WHAT THEY BRING — EACH WITH ITS OWN STANDARD

- **Data** — ISO 19115, GeoDCAT-AP
- **Software** — CodeMeta
- **Publications** — DataCite
- **Services** — EOSC service profile

THE CHALLENGE

How can a Node share its resources without losing information, while still meeting EOSC Federation requirements?

Semantic interoperability by design

Following the EOSC Semantic Interoperability Task Force recommendations — and going further

TF recommendation	Priority	How the Data Terra catalogue implements it
2.2.1 Latest DCAT as baseline, reuse application profiles	**Critical**	DCAT 3 pivot for every resource, aligned with DCAT-AP and GeoDCAT-AP
2.2.2 Standard domain vocabularies and identifier schemes	**Critical**	Values are SKOS concepts referenced by IRI; ROR, ORCID, SPDX, typed PIDs
2.2.3 Dedicated artefacts for outputs other than data	**Important**	CodeMeta for software, DataCite for publications, MOD for semantic artefacts
2.2.5 Modelling and validation alongside the profile	**Useful**	SHACL conformance gate before anything is exposed
2.4.1 Semantic artefacts managed through a catalogue	**Critical**	Disciplinary vocabularies resolved through EarthPortal, the Earth-science SAC

…and further: the node keeps the full, rich description and projects a compliant view to the EOSC EU Node.

Source: EOSC-A Semantic Interoperability TF, Deliverable 2 — Practical semantic interoperability recommendations for the EOSC Federation (2026)

One DCAT 3 pivot for every resource

Native descriptions in, one knowledge graph, several standard projections out

NATIVE SOURCES

- ISO 19115 / GeoDCAT-AP
- DataCite
- CodeMeta
- MOD
- EOSC service profile

DCAT 3 PIVOT

dcat:Resource + dct:type

- Common core shared by all types
- Thin per-type extension borrowed from the community standard
- One mapping per source, property by property

PROJECTIONS

EEN · research products

OAI-PMH, OpenAIRE / DataCite profile

EEN · services

EOSC JSON service profile

Peer nodes

DCAT JSON-LD via the node API

COMMON CORE

identity · title & description · dates · publisher, creators, contacts · licence & access rights · themes & keywords · spatial & temporal coverage · distributions · typed identifiers (adms:identifier) · equivalences (owl:sameAs) · provenance (PROV-O)

A typed dcat:Resource, not nine subclasses

Plain DCAT 3, with the kind of resource carried by a SKOS concept

DCAT 3 only has classes for datasets, data services and catalogues. Forcing software, publications or other research product into dcat:Dataset would state something false.

So every record is a **dcat:Resource**, and its kind is a concept in **dct:type**

No custom application profile to maintain; the type vocabulary can evolve without touching the model.

```
# Dataset ·
<https://w3id.org/earth-data/resource/4c6312e0-…>
  a dcat:Resource, dcat:Dataset ;
  dct:type coar:c_ddb1 ;
  dct:title "Chemical and mineralogical composition
     of southern Africa dust aerosols" ;
  dcat:theme <terra-vocabulary.org/…/c_4064c319> ;
  dct:license spdx:CC-BY-4.0 ;
  dct:accessRights coar:c_16ec ;
  adms:identifier [ dct:type "DOI" ;
     skos:notation "10.57932/b4373977-…" ] ;
  dct:source <terra-vocabulary.org/id/DataSet/640a…> ;
  prov:hadPrimarySource <doi.org/10.57932/b4373977-…> .
```

```
# Publication ·
<https://w3id.org/earth-data/resource/710e0264-…>
  a dcat:Resource ;
  dct:type coar:c_6501 ;   # journal article
  dct:title "A real scale application of a novel set of
     spatial and similarity features … from DAS data" ;
  bibo:journal "Geophysical Journal International" ;
  dct:license <creativecommons.org/licenses/by/4.0/> ;
  dct:accessRights coar:c_abf2 ;
  adms:identifier [ dct:type "DOI" ; skos:notation
     "10.1093/gji/ggae382" ], [ dct:type "HAL" ;
     skos:notation "insu-04763870" ] ;
  owl:sameAs <doi.org/10.1093/gji/ggae382> ;
  prov:hadPrimarySource <insu.hal.science/insu-04763870v1> .
```

NINE RESOURCE TYPES — ONE MODEL

Dataset | **Research software** | **Publication** | **Service** | **Semantic artefact**

Project | **Training material** | **Deployable application** | **Interoperability guideline**

Values are concepts, not strings

Controlled vocabularies referenced by IRI, for every resource type

- `dct:type` — **COAR Resource Types**
- `dcat:theme (data)` — **Data Terra Thesaurus**
- `earthdata:scientificDomain` — **EuroSciVoc**
- `Service category` — **EOSC Service Classification**
- `Service TRL, access type` — **EOSC vocabularies**
- `dct:license` — **SPDX licence list**
- `dct:accessRights` — **COAR Access Rights**
- `dct:publisher, funders` — **ROR**
- `dct:creator, contributor` — **ORCID**
- `dct:language` — **ISO 639**
- `Related identifiers` — **DataCite RelationType**
- `Coordinate Reference System` — **OGC EPSG**

Why it matters A concept IRI carries its label, definition and alignments: a node that receives the record can filter, translate and reason on it — a free-text string only gets matched character by character.

Identity and history of every record

Persistent identifiers on w3id, provenance with PROV-O

"PERSISTENT" IDENTIFIERS

w3id.org/earth-data/…

- **Node-minted, stable, resolvable**
  Every record gets its own IRI, independent of the source system
- **Native PIDs kept, typed**
  DOI, SWHID, HAL-ID, Handle as adms:identifier with their scheme agency
- **Source identity preserved**
  Original IRIs as owl:sameAs; federated records keep their node's identifier

PROVENANCE WITH PROV-O

Why: every harvested record keeps a machine-readable link to its source, so it can be traced back, credited and re-harvested.

- **Link to the source**
  prov:hadPrimarySource → the original record
- **Credit the producers**
  prov:wasAttributedTo → authors and organisations, with their role when known
- **Trace our processing**
  prov:wasGeneratedBy → the "DataTerra harvest" activity and each enrichment step

Queryable like any other metadata, and carried over into the EOSC exports.

Rich inside, compliant outside

EOSC EU Node (EEN) compatibility is a projection of a richer model

Node knowledge graph

full DCAT 3 description, all sources, all types

→

Validation control

SHACL shapes on mandatory EEN fields: incomplete records are blocked and flagged; PID type and licence are reported

→

EEN · research products OAI-PMH, OpenAIRE / DataCite

EEN · services EOSC JSON service profile

BETWEEN NODES: WHY A RICH DCAT 3 JSON-LD OUTPUT IS NEEDED

EEN formats — mandatory, but not sufficiently expressive

- **Fixed field set per channel** (OpenAIRE / DataCite, EOSC profile)
- **Domain extensions dropped**: GeoDCAT-AP, CodeMeta, MOD
- **Concepts reduced to labels, provenance reduced to names**

→

DCAT 3 JSON-LD — node to node, without reduction

- **The full pivot record**, as stored in the node
- **Vocabularies, identifiers, provenance** kept as IRIs
- **Peers align, deduplicate and enrich** — no loss

EOSC Federation: how nodes find and read each other

Federation · discovery through /.well-known, tested with the LSCN node

DISCOVERY ·
https://resources.earth-data.eu/.well-known/eosc-federation/node

Why it is needed In a federation without a central hub, a node must find the others' catalogues, formats and support services by itself — not through URLs exchanged by e-mail and configured by hand.

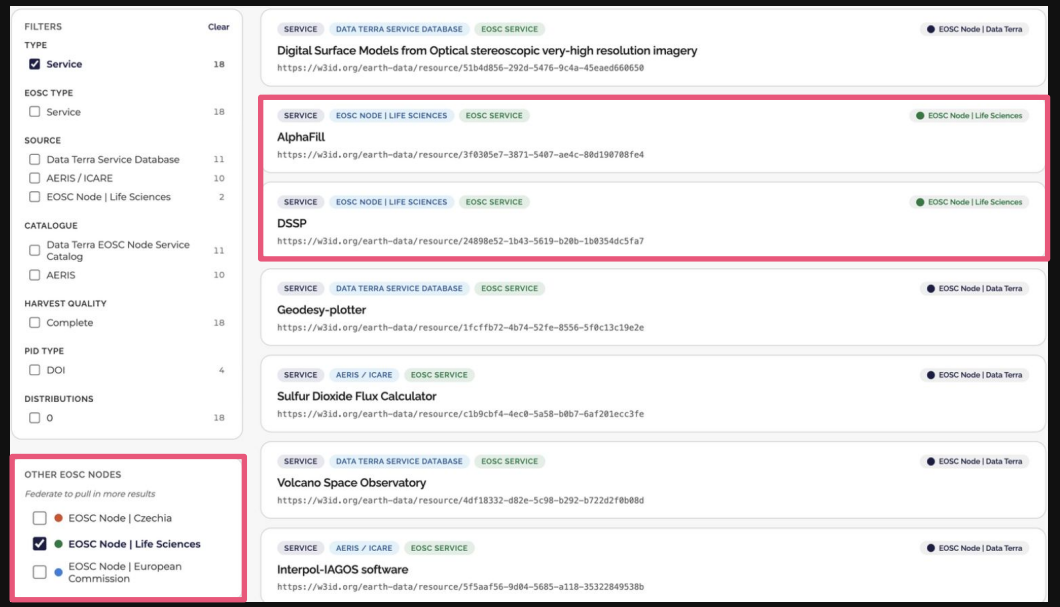

```
node_endpoint:   "https://resources.earth-data.eu/.well-known/eosc-federation/node"
capabilities:
  0:  { capability_type: "Resource Catalogue", endpoint: "https://resources.earth-data.eu/api/resources?eoscTypes=Research%20Product&format=dcat",
  1:  { capability_type: "Resource Catalogue", endpoint: "https://resources.earth-data.eu/oai", version: "0.2.56", … }
  2:  { capability_type: "Service Catalogue", endpoint: "https://services.earth-data.eu/api/resources?types=Service&format=dcat", version: "0.2.56'
  3:  { capability_type: "Service Catalogue", endpoint: "https://services.earth-data.eu/services", version: "0.2.56", … }
  4:  { capability_type: "AAI", endpoint: "https://sso.earth-data.fr/realms/gaia-data/.well-known/openid-configuration", protocol: "OIDC", … }
  5:  { capability_type: "Monitoring", status: "PLANNED" }
  6:  { capability_type: "Accounting", status: "PLANNED" }
  7:  { capability_type: "Helpdesk", status: "PLANNED" }
  8:  { capability_type: "Order Management", status: "PLANNED" }
  9:  { capability_type: "Application Workflow Management", status: "PLANNED" }
node_details:  { name: "Data Terra EOSC Node", url: "https://www.earth-data.eu", description: "The Data Terra EOSC Node." }
```


WHAT IT BRINGS

- **One predictable URL per node** — all a peer needs to know
- **Machine-readable capabilities** — endpoints, formats, status
- **Self-configuring harvesters** — no bilateral set-up

Test with the Life Science Connect node

23 of our services read by LSCN · **10** LSCN services ingested



Our catalogue: LSCN services (green badge) among Data Terra ones

- **Found via descriptor**, 0 warnings
- **DCAT ≈ 1:1** onto pivot
- **EOSC Beyond** format agreed
- **Open:** identity & provenance

eosc Testing node-to-node exchange with ENVRI

DEDUPLICATION MECHANISM

DATASET · EOSC RESEARCH PRODUCT · 5 RECORDS DUPLICATED · CROSS-CATALOGUE

Argo float data and metadata from Global Data Assembly Centre (Argo GDAC)

MTEP	https://w3id.org/earth-data/resource/1b2a81fd-3eff-5129-a5ce-eed32b17fd98
ENVRI-HUB NEXT	https://w3id.org/earth-data/resource/df502b0b-1822-53e7-bb2e-5b28c0ffd475
ENVRI-HUB NEXT	https://w3id.org/earth-data/resource/36ac37c3-871d-55ae-a14f-76e54b8086b2
ENVRI-HUB NEXT	https://w3id.org/earth-data/resource/b2b6b718-d339-5ba9-8bcb-82118b6638bb
ENVRI-HUB NEXT	https://w3id.org/earth-data/resource/015304ea-68b3-58d7-ae6a-906df6e935c7

DATASET · EOSC RESEARCH PRODUCT · 2 RECORDS DUPLICATED · CROSS-CATALOGUE

IAGOS Daily mean profiles at 4 levels - L3

AERIS	https://w3id.org/earth-data/resource/3751c1fd-7628-5c56-9ad8-71758983ee87
ENVRI-HUB NEXT	https://w3id.org/earth-data/resource/c4611458-21a9-5f7f-8a27-f3f125980968

DATASET · EOSC RESEARCH PRODUCT · 2 RECORDS DUPLICATED · CROSS-CATALOGUE

IAGOS Daily aggregated regional data at 3 levels - L3

AERIS	https://w3id.org/earth-data/resource/b566b470-79e1-5d27-8b11-1c0f1f3d3a79
ENVRI-HUB NEXT	https://w3id.org/earth-data/resource/cec0ef85-ff63-5d52-9c7b-2ab5088a55e0

128	72	56	0
ENVRI records indexed	datasets	services	errors

After deduplication 56 datasets + 55 services visible

MECHANISM — GROUP, NEVER MERGE OR DELETE

1. **Keep identity** own w3id @id; source ID and URI kept
2. **Match** same DOI / PID, or title + declared type
3. **Cluster** shared clusterId + confidence level
4. **Keep all** one shown, every member kept

WHAT THE CLUSTERS SAY

- **IAGOS × 2 — true duplicates**
  same object in AERIS and ENVRI (DOI 10.25326/638, /639)
- **Argo — strong title match**
  but granularity issue

Lesson shared DOIs make true duplicates obvious; title matches expose the granularity question. WORK in progress, not yet in production.

Should the Federation converge on a DCAT based schema?

Layer 1 · common floor

EEN profile for every node — mandatory, minimal, central discovery

Layer 2 · rich contract

DCAT 3 pivot + shared vocabularies between neighbouring and thematic nodes that already use enriched models

1 **Rich and compliant is possible**

one DCAT 3 pivot, several standard projections

2 **Identifiers and granularity are the hard part**

more than syntax or serialisation

3 **Next: test with more nodes**

DTO, national nodes, …

"There is a consensus in the nodes that exposing rich domain catalogue resources would be necessary to maximise the benefits of federating EOSC node catalogue resources." — EOSC United, node interviews

Thank you · christelle.pierkot@cnrs.fr, julien.homo@foxcub.fr, alessandro.rizzo@ird.fr