

PUBMET2025 - Infrastructure and Tools for Open Science

Federating Scientific Communities for Open Discovery: The LUMEN Data Mesh Framework

12 | 09 | 2025 by Julien Homo (Foxcub)



**Funded by
the European Union**

LUMEN is funded by the European Union under Grant Agreement no. 101187940. 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 Commission. Neither the European Union nor the granting authority can be held responsible for them.



LUMEN - In a nutshell

A federated ecosystem to empower interdisciplinary data discovery and reuse within EOSC



19 partners and affiliated entities



From 9 EU countries (FR, BE, IT, DE, PT, AT, PL, SE, HR) and the UK



About 70 people involved



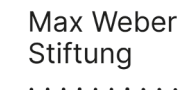
36 months project: 1st January 2025 - 31 December 2027



Project coordinator: CNRS



Budget: 7 million euros



LUMEN objectives and key contributions

Four strategic objectives to transform how communities discover, share, and reuse research outputs

Objectives

S01

Create a shared and collaborative cross-disciplinary knowledge and **data exchange space** for the next generation EOSC

S02

Enhance the qualitative discovery of research outputs by establishing an ecosystem of **innovative features**

S03

Support the production of **FAIR-by-design research outputs**

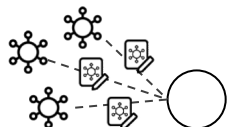
S04

Facilitate early stage of the research lifecycle by **enhancing specific tools and services**

Key contributions to EOSC



White-label discovery platforms



Data Mesh architectural framework



LUMIS: LUMEN Infrastructure for Semantics



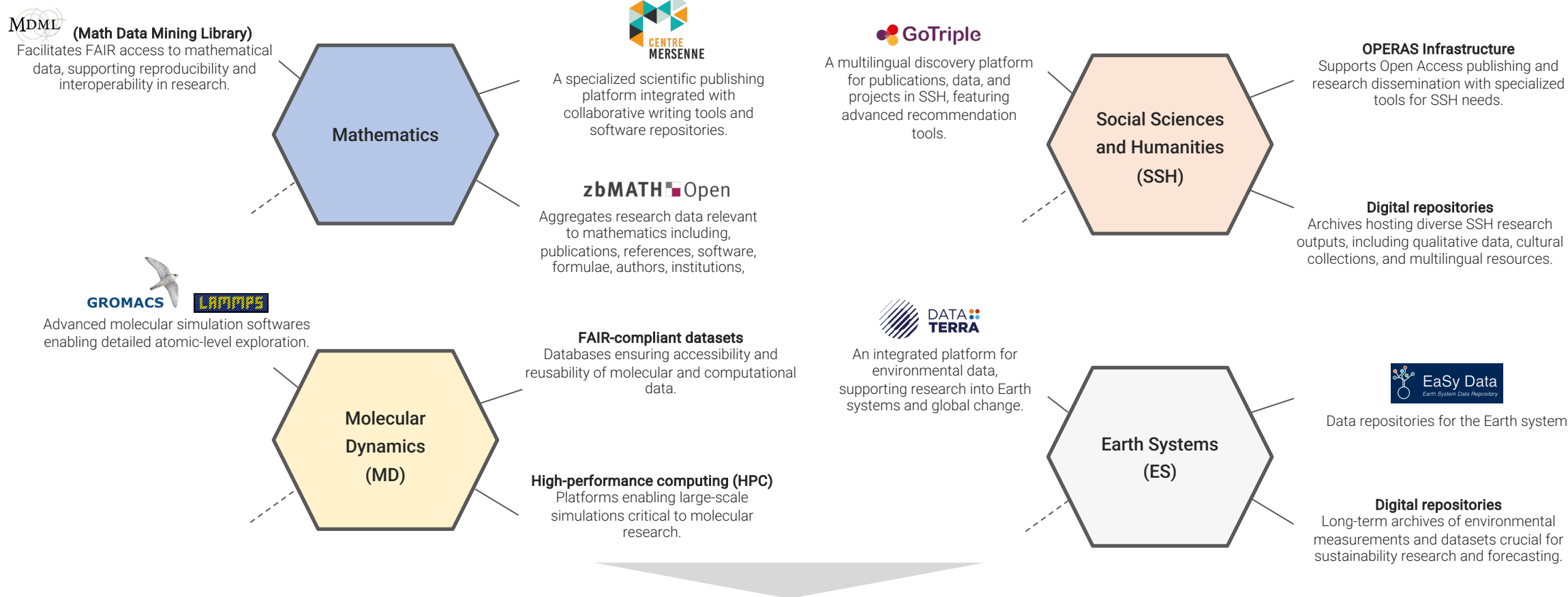
AI-enhanced services



Progressive integration into the EOSC Federation

4 scientific domains at the heart of LUMEN

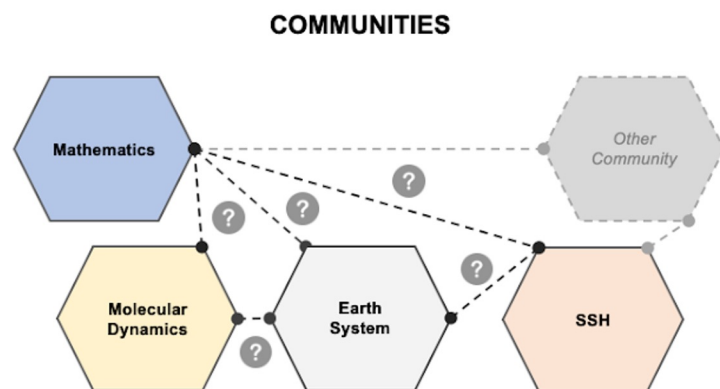
LUMEN is a federated collaboration of these scientific communities, leveraging their unique strengths, infrastructures, and expertise.



Organised through a data mesh

The LUMEN data mesh: enabling interdisciplinary collaboration

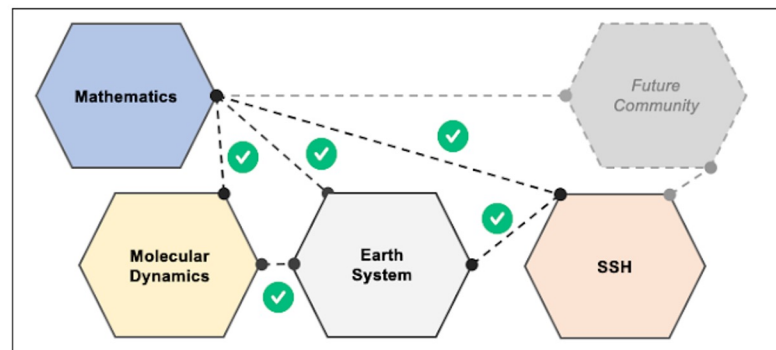
A three-layer framework to interconnect autonomous communities, align practices, and support reusable, FAIR-by-design scientific outputs.



From disconnected systems to a shared ecosystem, enabling **collaboration without centralisation.**

FEDERATED DATA SHARING GOVERNANCE

CONNECTED COMMUNITIES ECOSYSTEM



COMMON DISCOVERY INFRASTRUCTURE

Domain-agnostic accelerators

Federated innovative services

Domain-specific components

Interoperability and EOSC integration

The LUMEN data mesh in action

Translating community needs into modular, federated components for discovery, governance, and interoperability

Interoperable Platforms

- Domain-specific discovery portals interconnected via shared data contracts and governance.

White-Label Discovery Platform

- A flexible, EOSC-ready solution adaptable to each domain, ensuring interoperability by design.

FAIR-by-Design Infrastructure

- Tools like LUMIS support the creation of semantic artefacts, aligned with EOSC standards.

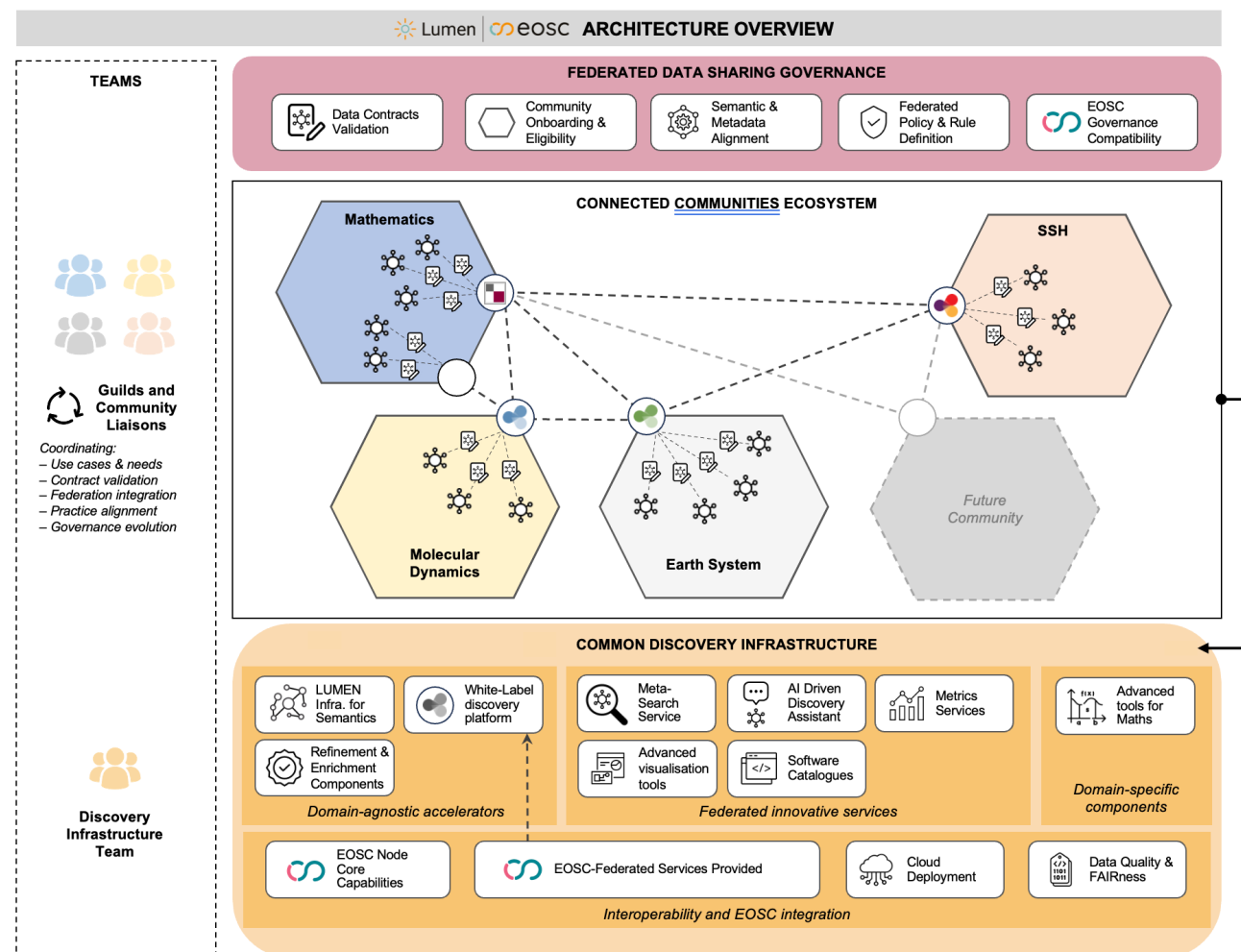
Cross-Domain Discovery

- Meta-search, AIDA, and visual tools enable exploration across datasets, disciplines, and services.

Composable & Community-Governed Services

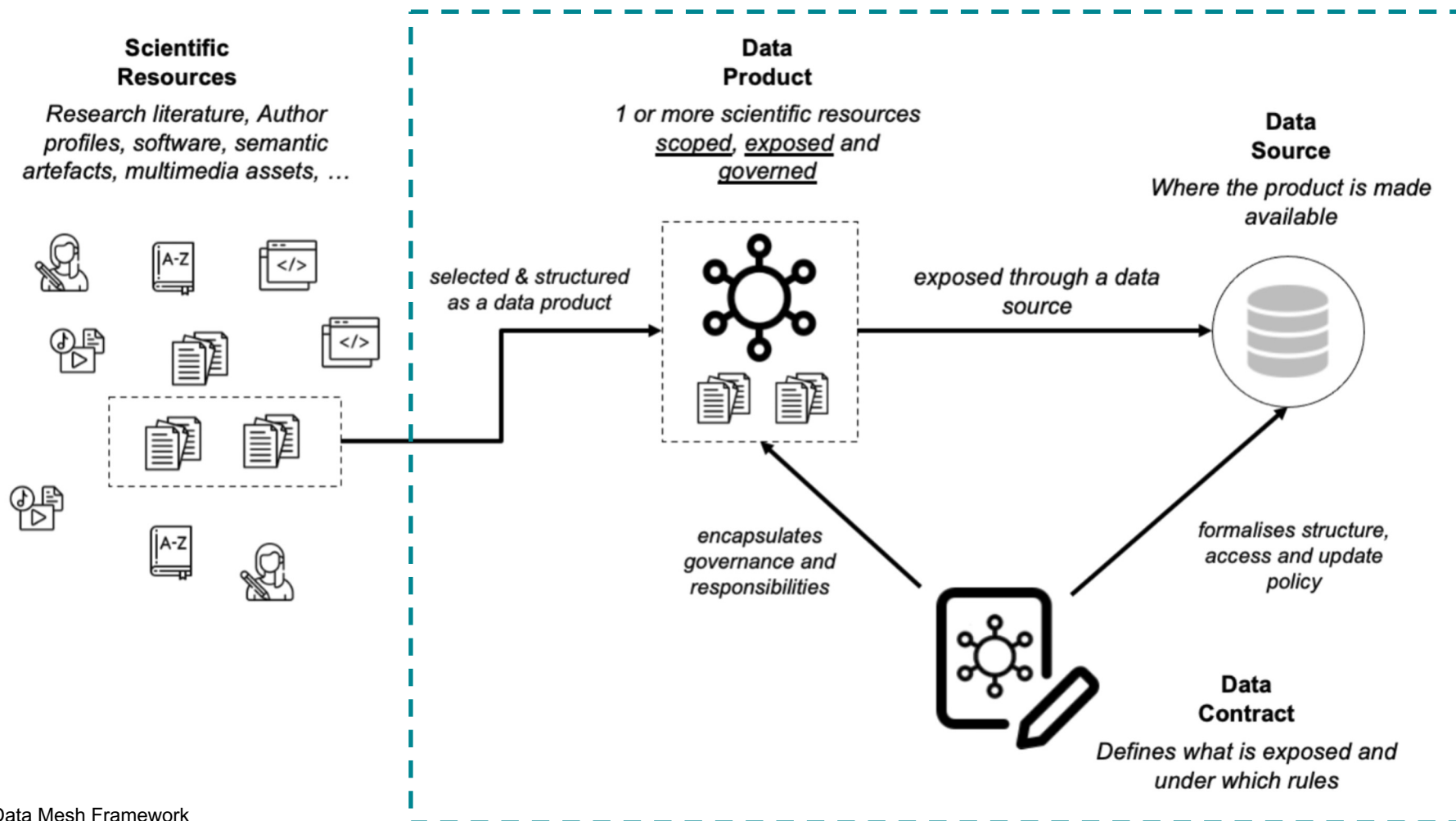
- Reusable catalogues and accelerators managed by domain guilds to foster ownership and reuse.

A modular, federated architecture rooted in community autonomy and interoperability



From scientific resources to governed data products

Understanding the key components that make data reusable, governable, and interoperable in the LUMEN Data Mesh.



LUMEN data
mesh
key concepts

Example of a LUMEN data contract

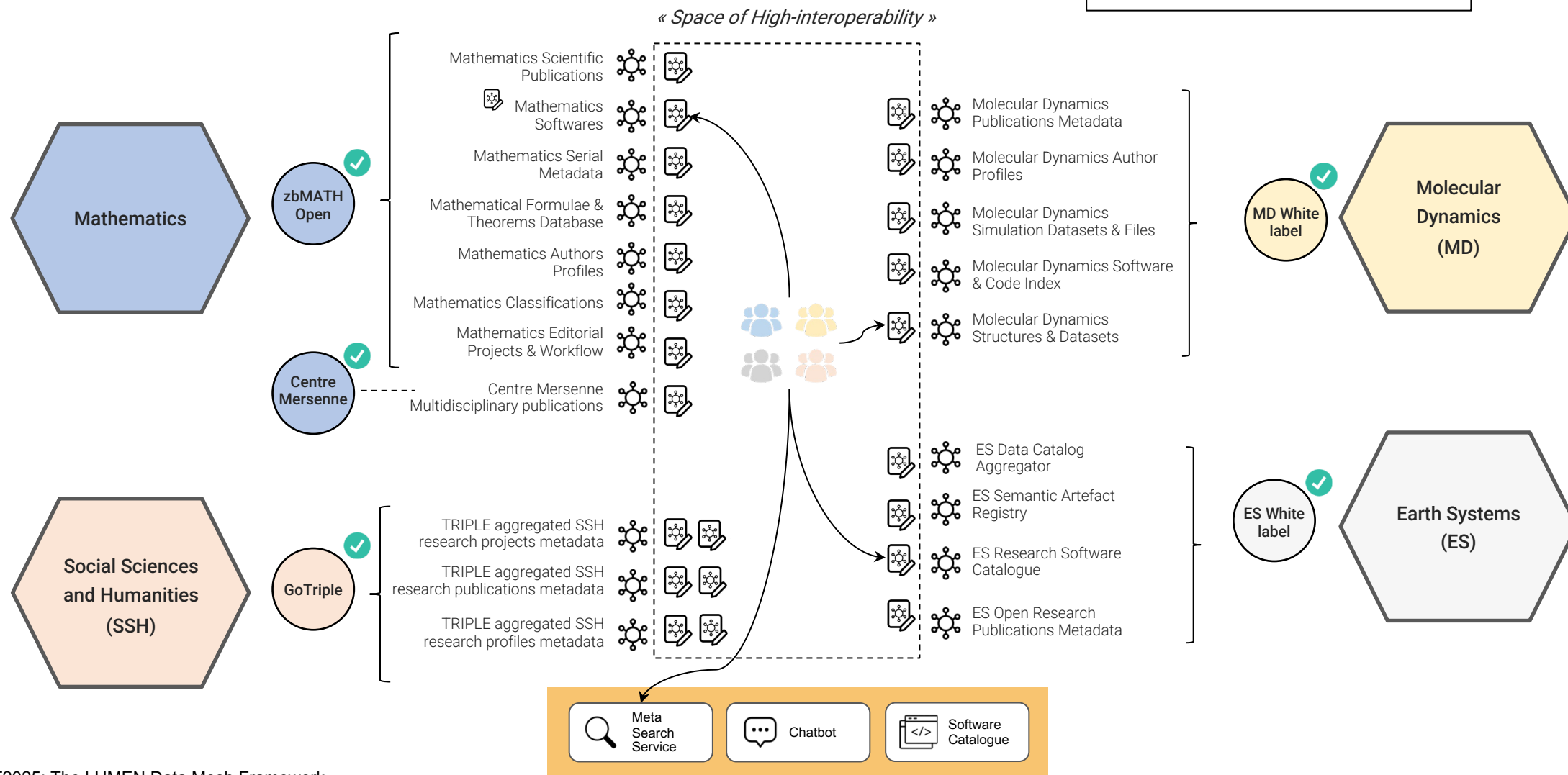
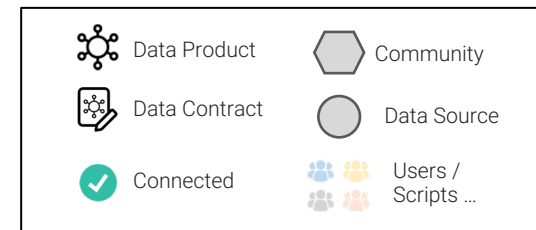


Mathematical
Publications – Centre
Mersenne and
Geodesic

```
id: contract-math-pubs-centre-mersenne
name: Mathematical Publications – Centre Mersenne and Geodesic
version: 1.0.0
type: producer
dataProduct:
  id: math-publications
  title: Open Access Mathematical Publications
  description: >
    A curated collection of open-access mathematical publications,
    including journal
    articles and conference proceedings, with structured metadata and
    links to full text
    in PDF and JATS-XML formats.
  resourceType: publicationCollection
  formats:
    - application/pdf
    - application/jats+xml
    - application/mathml+xml
  keywords:
    - Mathematics
    - Scientific Publications
    - JATS
    - Open Access
    - Scholarly Communication
  license: https://creativecommons.org/licenses/by/4.0/
  version: continuous
dataSource:
  - id: centre-mersenne
    title: Centre Mersenne
    curator: Mathdoc / CNRS
    endpoint: https://www.centre-mersenne.org/api/
    protocol: OAI-PMH
  - id: geodesic-mdml
    title: Geodesic MDML Catalogue
    curator: Geodesic Initiative
    endpoint: https://geodesic.math/api/
    protocol: REST+JSON
    status: in-development
```

```
roles:
  - actor: Centre Mersenne team
    role: producer
    responsibilities:
      - Metadata production
      - Full-text hosting
      - Semantic structuring (JATS-XML, MathML)
      - Update and quality assurance
  - actor: Geodesic MDML team
    role: producer
    responsibilities:
      - Metadata ingestion
      - Long-term preservation planning
      - API development and exposure
updatePolicy:
  frequency: monthly
  retentionPolicy: indefinite
  versioning: supported
  metadataValidation: automated + human review
quality:
  semanticStructure: JATS-XML compliant
  metadataCompleteness: high
  accessConsistency: guaranteed
access:
  visibility: public
  downloadAllowed: true
  usagePolicy: open-access / citation-required
notes:
  - >
    This producer contract reflects full end-to-end governance by the
    mathematics
    community, covering metadata, full-text formats, quality, and platform
    availability.
  - >
    The data product is eligible for federation in LUMEN based on strong
    FAIR alignment
    and machine-actionable exposure.
```

LUMEN data products & data contracts



From design to practice what we're beginning to see

LUMEN is still in early implementation, but concrete design challenges and structuring choices are already emerging



Identified challenges

- Bridging heterogeneous maturity, infrastructures, and governance
- Operationalising FAIR in distributed, domain-driven ecosystems
- Designing a mesh where benefits appear late
- Engaging established communities in future cross-domain value
- Sustaining alignment while respecting sovereignty

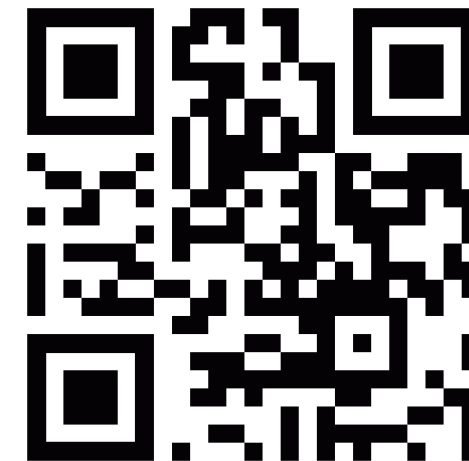


Key Observations

- Success depends more on shared value than technical integration
- Cross-domain use cases are the strongest proof of value, but take time
- Autonomy enables trust and long-term alignment
- Democratization requires time, capacity building, shared language, and trust

Hvala!

Please feel free to ask any questions.



And don't hesitate to reach out at julien.homo@foxcub.fr



**Funded by
the European Union**

LUMEN is funded by the European Union under Grant Agreement no. 101187940.
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 Commission. Neither
the European Union nor the granting authority can be held responsible for them.

