

Les plateformes de découverte du projet LUMEN : un enjeu majeur d'interdisciplinarité et de promotion de la science

OS DAYS - GRICAD
26 Novembre 2025
Suzanne Dumouchel



**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.



eosc LUMEN - En quelques indicateurs

Un écosystème fédéré pour favoriser la découverte et la réutilisation interdisciplinaires des données au sein de l'EOSC



19 partenaires et entités affiliées



9 pays EU (FR, BE, IT, DE, PT, AT, PL, SE, HR) et UK



Environ 70 personnes impliquées



Projet de 36 mois : 1 Janvier 2025 - 31 Decembre 2027



Coordinateur du projet: CNRS



Budget: 7 million euros



OBJECTIFS ET CONTRIBUTIONS PRINCIPALES DE LUMEN

Quatre objectifs stratégiques pour transformer la façon de découvrir, de partager et de réutiliser les ressources scientifiques pour les communautés scientifiques

Objectifs

S01

Créer un espace partagé et collaboratif d'échange de connaissances et de données interdisciplinaires pour la prochaine génération d'EOSC.

S02

Améliorer la découverte qualitative des résultats de recherche en mettant en place un écosystème de fonctionnalités innovantes.

S03

Soutenir la production de résultats de recherche FAIR dès leur conception

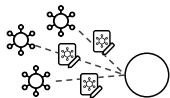
S04

Faciliter les premières étapes du cycle de vie de la recherche en améliorant certains outils et services spécifiques.

Contributions majeures à la Fédération EOSC



White-label
discovery platforms



Data Mesh
architectural
framework



LUMIS: LUMEN
Infrastructure for
Semantics



AI-enhanced
services

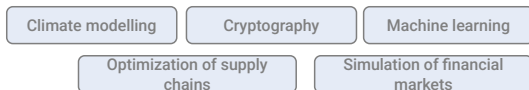


Progressive
integration into the
EOSC Federation

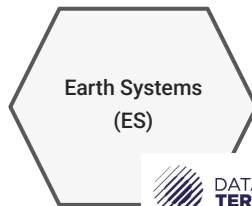
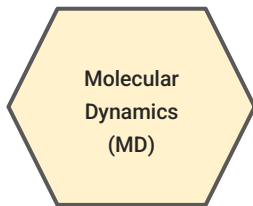
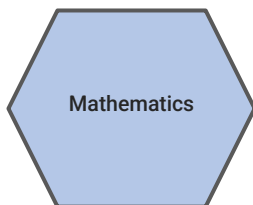
4 DOMAINES SCIENTIFIQUES AU COEUR DE LUMEN

LUMEN est une collaboration fédérée entre ces communautés scientifiques, qui tire parti de leurs atouts, infrastructures et expertises uniques..

Une discipline fondamentale qui fournit des méthodologies et des outils rigoureux pour résoudre des problèmes complexes.



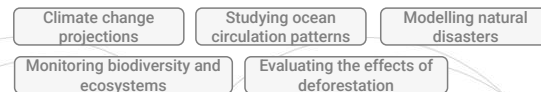
Une branche des sciences physiques qui utilise des simulations numériques pour prédire les interactions moléculaires et atomiques..



Axé sur l'exploration des dimensions humaines des sociétés, notamment la culture, la langue et l'organisation sociale.



Une étude intégrée des interactions entre les composantes de la Terre afin de relever les défis environnementaux mondiaux.



Organisés autour d'un data mesh

IMPACTS ET LIVRABLES DE LUMEN

Révolutionner la recherche interdisciplinaire grâce à des plateformes FAIR basées sur l'IA afin de stimuler la découverte, la collaboration et la confiance.

Impact 1: Faciliter l'accès à des données multidisciplinaires de haute qualité, conformes aux principes FAIR, afin de soutenir la recherche interdisciplinaire.



- Une plateforme de découverte en marque blanche offrant des services prêts à l'emploi et personnalisables pour les communautés
- Une fédération de 4 plateformes spécifiques à un domaine afin de promouvoir l'interopérabilité et la réutilisation interdisciplinaire

Impact 2: Contribuer à la Fédération en fournissant des services réutilisables et conformes à l'EOSC via les nœuds EOSC, afin d'améliorer la découvrabilité et l'interopérabilité sémantique.



- Plateforme de découverte en marque blanche
- Infrastructure LUMEN pour la sémantique (LUMIS)

Impact 3: Améliorer l'efficacité de la recherche et l'accès à l'expertise grâce à des outils évolutifs et éducatifs.



- Assistant de découverte alimenté par l'IA (AIDA) permettant une exploration interactive en langage naturel

Impact 4: Révolutionner le processus de recherche en améliorant l'expérience utilisateur, la reproductibilité et la collaboration.



- Outils multi-plateformes pour la découverte : méta-recherche, chatbot, métriques, visualisation et catalogue de logiciels

Impact 5: Transformer la gouvernance des données en favorisant l'autonomie, l'interopérabilité et l'alignement stratégique entre les communautés.



- Architecture de maillage de données reliant des domaines de données autonomes
- Modèle de gouvernance fédérée pour le partage de données piloté par la communauté

Défis identifiés



- Concilier des niveaux hétérogènes de maturité, d'infrastructures et de cultures de gouvernance entre les domaines.
- Mettre en œuvre les principes FAIR au sein d'écosystèmes distribués et axés sur les domaines.
- Concevoir un cadre maillé dont les avantages apparaissent tardivement, ce qui rend l'adhésion initiale plus difficile.
- Convaincre les communautés établies de s'engager dans un cadre dont la valeur réside dans la convergence future entre les domaines.
- Maintenir l'alignement au fil du temps tout en respectant la souveraineté des communautés.



Observations

- Le succès d'un maillage de données dépend moins de l'intégration technique que de la capacité à projeter une valeur partagée entre les domaines.
- Les utilisations interdomaines sont la démonstration la plus puissante de la valeur, mais elles apparaissent tardivement dans le cycle de vie.
- L'autonomie n'est pas un obstacle à la fédération, c'est la condition préalable à la confiance et à l'alignement à long terme.
- Les communautés ont besoin d'espace pour expérimenter avant de s'engager dans une intégration complète.



LUMEN jette les bases d'un EOSC fédéré et géré par la communauté. Son architecture préserve l'autonomie et la capacité d'action des communautés de recherche, tout en permettant une interopérabilité à long terme et une synergie entre les domaines grâce à une infrastructure axée sur la sémantique et conçue selon les principes FAIR.



Recommandations à l'intention des pairs et des acteurs de l'EOSC

- Reconnaître que l'adoption est un processus de confiance et de projection, et pas seulement de conformité.
- Donner la priorité à la co-conception de cas d'utilisation afin de rendre le maillage significatif dès le premier jour.
- Soutenir les parcours d'intégration progressifs : le maillage ne nécessite pas une participation « tout ou rien ».
- Combiner les éléments techniques de base avec la facilitation communautaire et les cadres de gouvernance.

LUMEN – Genèse d'un écosystème fédéré

S'appuyer sur GoTriple pour étendre la découverte à de nouveaux domaines scientifiques, organisés selon une approche de maillage de données.

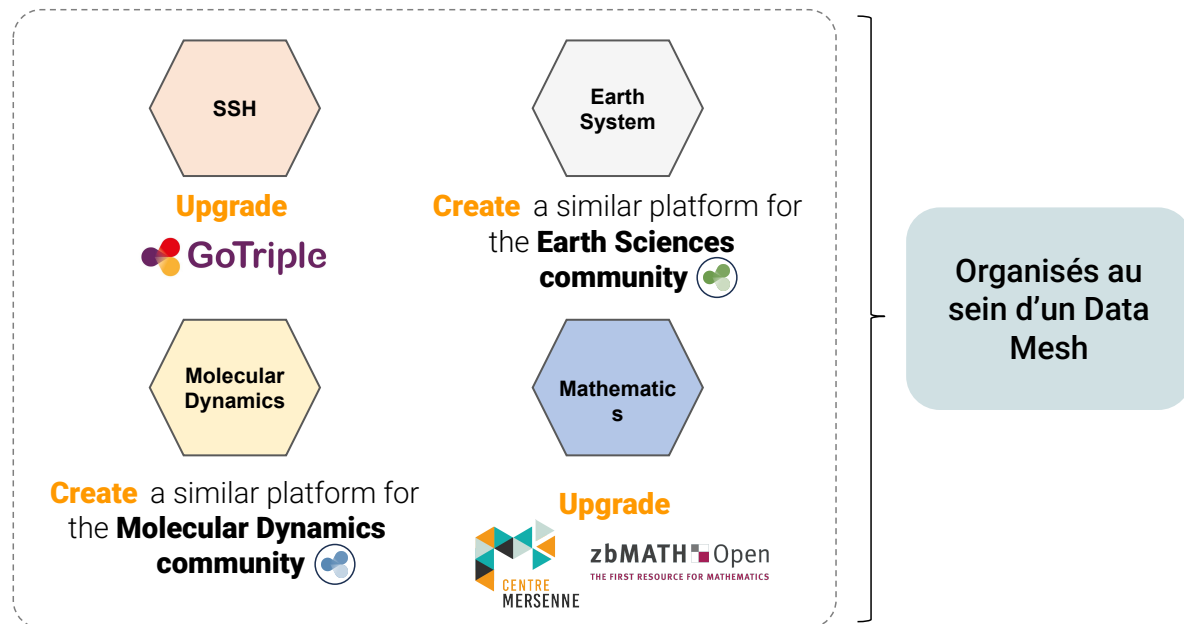


Plateforme de découverte multilingue
pour les Sciences Humaines et
Sociales

EU - project: 2019-2023

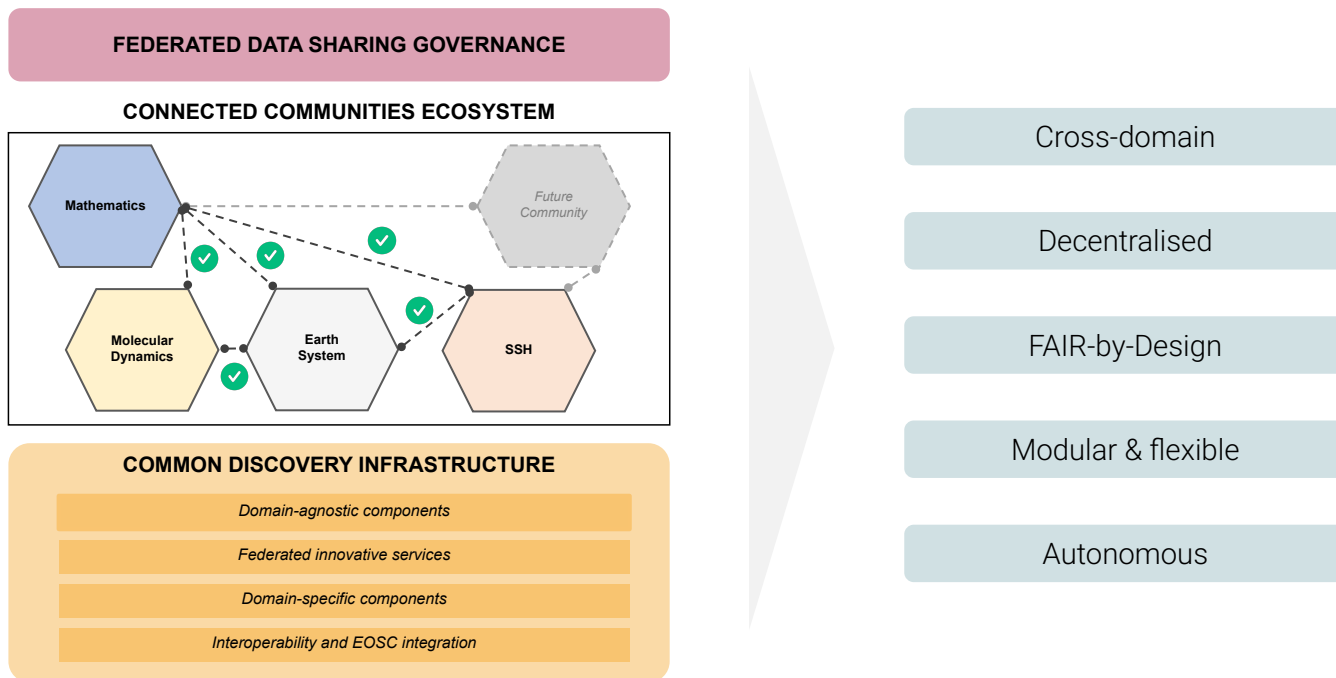


EU - project: 2025-2027



LUMEN data mesh for interdisciplinarity

Connecting autonomous communities through a federated and FAIR-by-design framework



LE DATA MESH EN ACTION

Traduire les besoins de la communauté en composants modulaires et fédérés pour la découverte, la gouvernance et l'interopérabilité

Plateformes interoperables

- Portails de découverte spécifiques à un domaine, interconnectés via des contrats de données partagées et une gouvernance commune.

La plateforme de découverte en marque blanche

- Une solution flexible, compatible avec l'EOSC, adaptable à chaque domaine et garantissant l'interopérabilité dès sa conception.

FAIR-by-Design Infrastructure

- Des outils tels que LUMIS facilitent la création d'artefacts sémantiques, conformes aux normes EOSC.

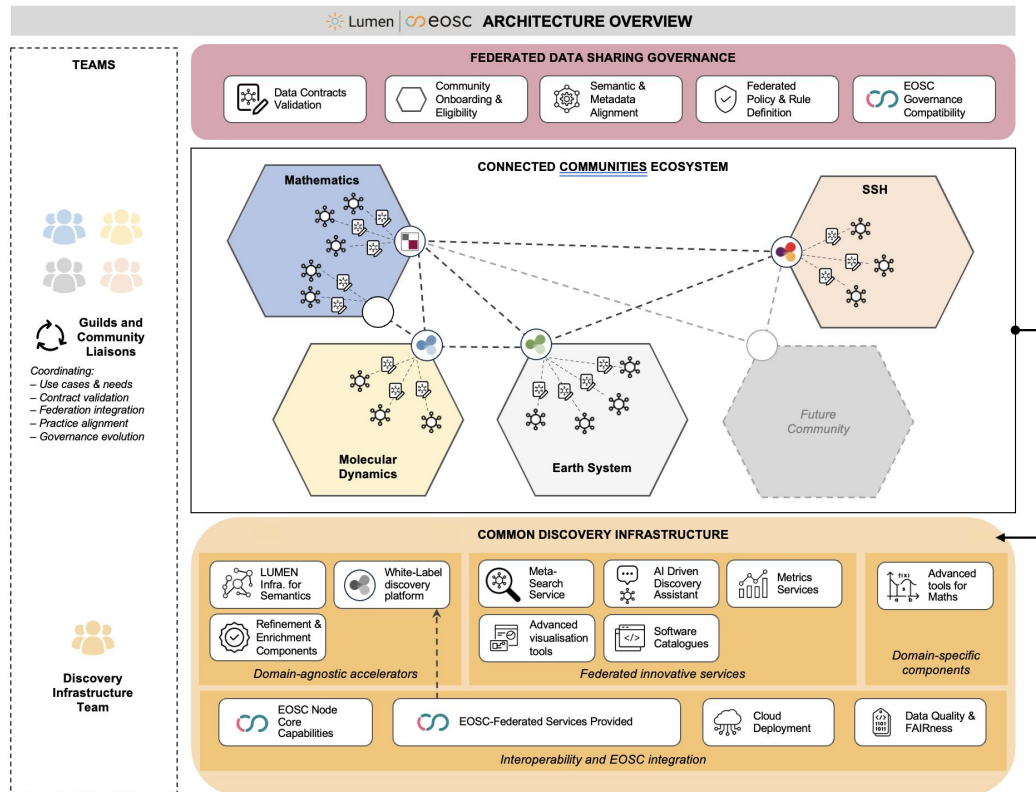
Cross-Domain Discovery

- La méta-recherche, l'AIDA et les outils visuels permettent d'explorer différents ensembles de données, disciplines et services.

Services modulables et gérés par la communauté

- Favoriser l'appropriation et la réutilisation par les communautés

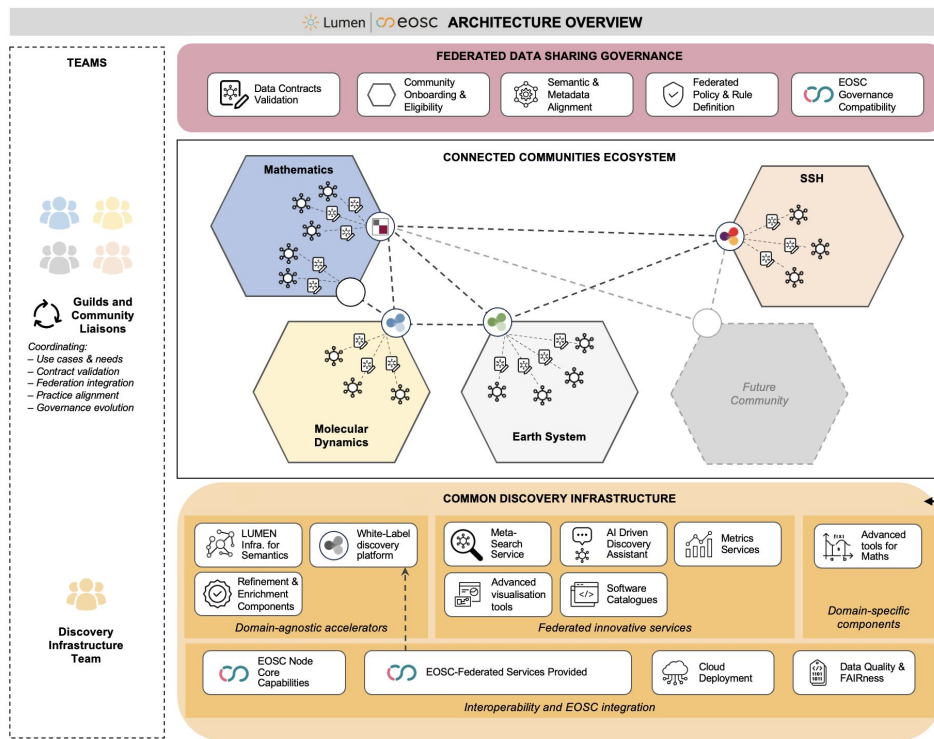
Une architecture modulaire et fédérée fondée sur l'autonomie communautaire et l'interopérabilité



LUMEN en 2 concepts clés

Un maillage de données fédéré reliant les communautés, alimenté par des outils de découverte partagés

Le Data Mesh:
un maillage de
données comme
espace d'échange de
données
entre les
communautés

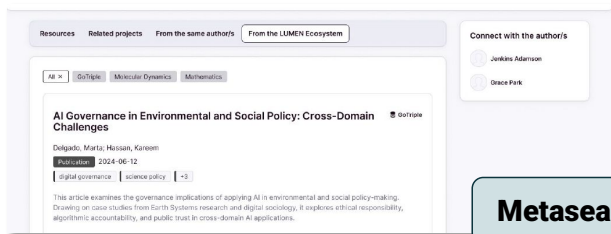


LUMEN Data Mesh
Architecture Framework, June 2025
<https://zenodo.org/records/15752126>

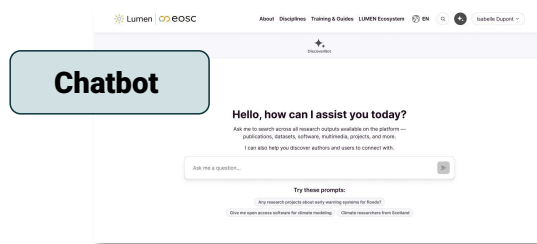
**8 outils innovants
pour faciliter la
découverte des
données**

8 outils innovants pour permettre la découverte des données

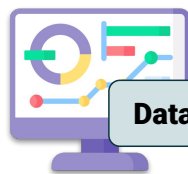
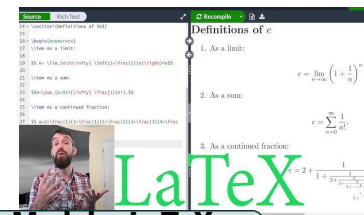
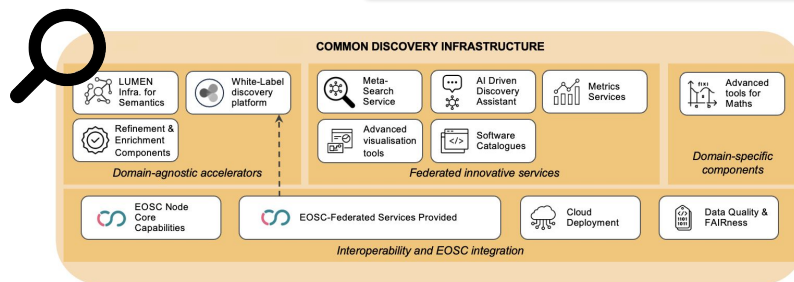
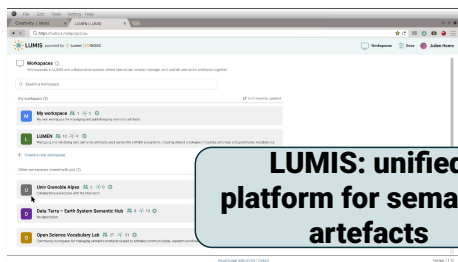
De l'alignement sémantique à l'exploration intelligente : une boîte à outils commune pour l'écosystème LUMEN



Metasearch

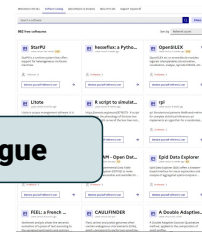


Chatbot



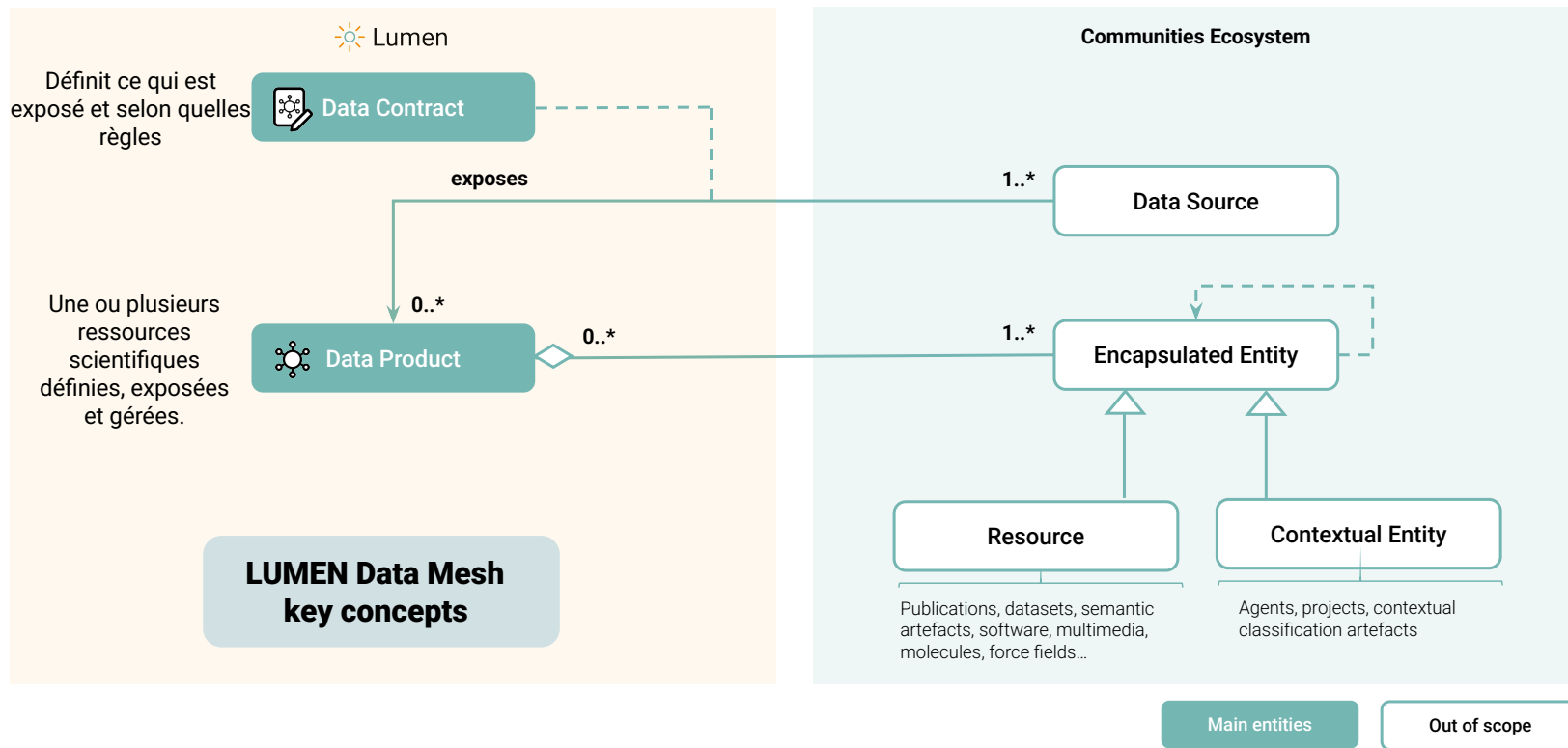
LUMEN Data Mesh Architecture Framework, June 2025
<https://zenodo.org/records/15752126>

Software catalogue



Des ressources scientifiques aux produits de données réglementés

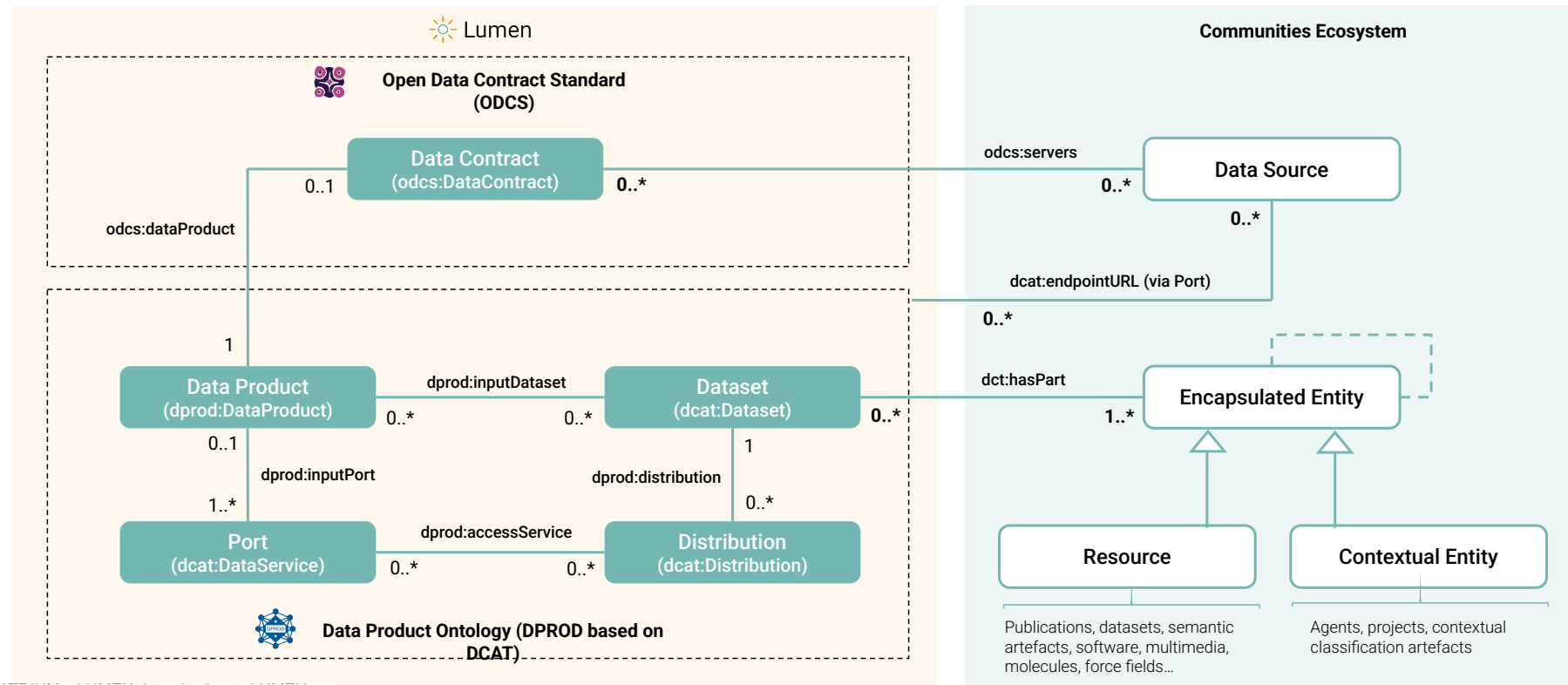
Comprendre les éléments clés qui rendent les données réutilisables, gérables et interopérables dans le Data Mesh LUMEN.



LUMEN Data Model

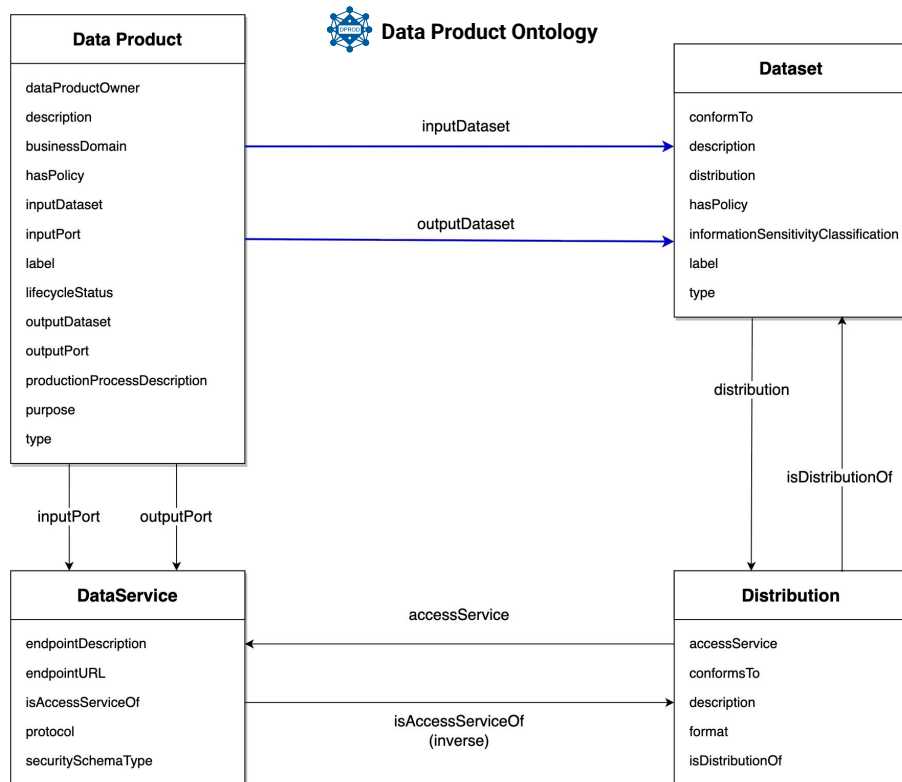
Main entities

Out of scope



Data Product Ontology (DPROD based on DCAT)

<https://ekgf.github.io/dprod>



DPROD Main aims

- To provide unambiguous and sharable semantics to answer the question: "What is a data product?"
- Be simple enough for any research community to adopt, yet expressive enough to support large-scale scientific data ecosystems
- Allow organisations to reuse their existing data catalogues and dataset infrastructure
- To share common semantics across different Data Products and promote harmonization

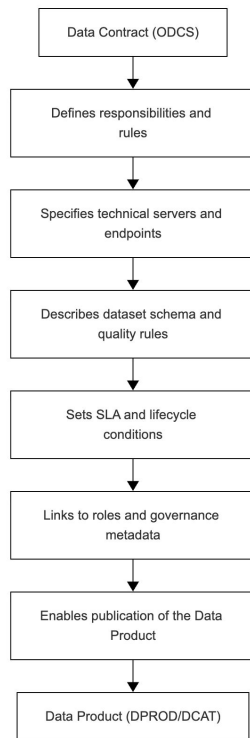
DPROD Profile

- **Data Mesh (dcat:Catalog)** - The collection of Data Products
- **Data Product (dprod:DataProduct)** - A data product may have input and output ports, code and metadata
- **Port (dcat:DataService)** - A digital interface that provides access to a Dataset. This can be an HTTP URL, a Database or a FileShare, etc.
- **Distribution (dcat:Distribution)** - A specific representation of a dataset (CSV, JSON, ADLS etc.) which can conform to a physical model
- **Dataset (dcat:Dataset)** - A collection of related data that can conform to a logical model

- It's one of the key advantages of the mesh: we don't aim to centralise or reinvent how each community structures its data.
- On the contrary, we let them retain full control over how they expose it — provided they clearly document their access points, formats, and the standards they rely on.

Open Data Contract Standard (ODCS)

<https://bitol-io.github.io/open-data-contract-standard/v3.0.2/>



Open Data Contract Standard (ODCS)

```

# What's this data contract about?
domain: seller # Domain
dataProduct: my quantum # Data product name
version: 1.1.0 # Version (follows semantic versioning)
status: active
id: 53581432-6c55-4ba2-a65f-72344a91553a

# Lots of information
description:
  purpose: Views built on top of the seller tables.
  limitations: Data based on seller perspective, no buyer information
  usage: Predict sales over time
  authoritativeDefinitions:
    - type: privacy-statement
      url: https://example.com/gdpr.pdf
  tenant: ClimateQuantumInc

kind: DataContract
apiVersion: v3.0.2 # Standard version (follows semantic versioning)

# Infrastructure & servers
servers:
  - server: my-postgres
    type: postgres
    host: localhost
    port: 5432
    database: pypl-edw
    schema: pp_access_views

# Dataset, schema and quality
schema:
  - name: tbl
    physicalName: tbl_1
    physicalType: table
    businessName: Core Payment Metrics
    description: Provides core payment metrics
    authoritativeDefinitions:

```

ODCS Main aims

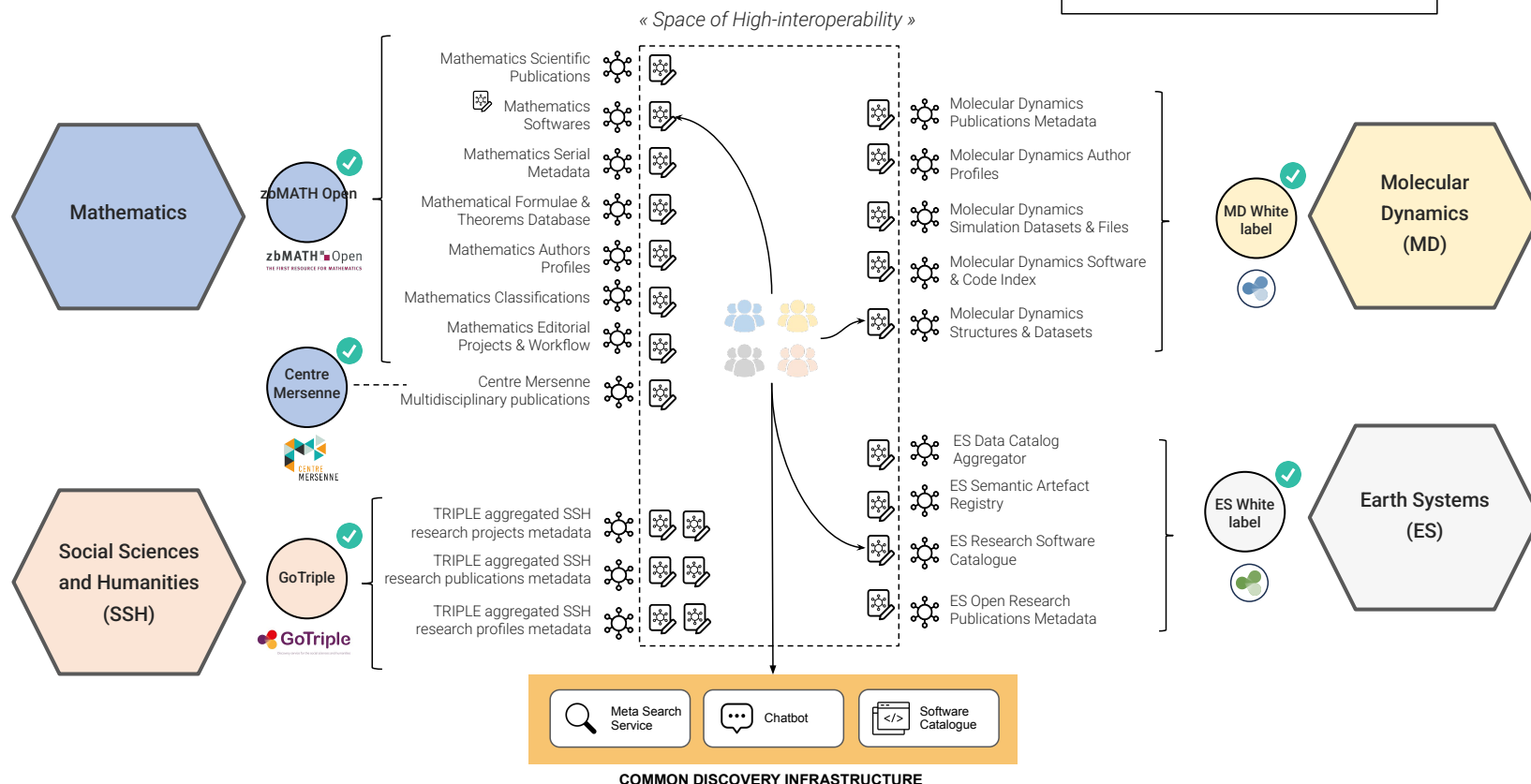
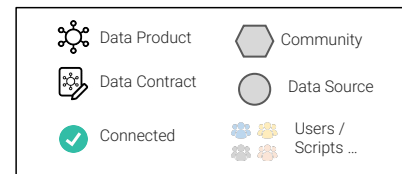
- To formalise data-sharing agreements in a structured, machine-actionable way.
- To make roles, responsibilities, and conditions of access explicit across federated systems.
- To enable automated validation, compliance, and lifecycle tracking of Data Products within the federation.
- To bridge human-readable policies (e.g. licences, attributions) with their automated enforcement through DPROD APIs (REST, OAI-PMH, SPARQL).

ODCS Profile

- Based on a modular, YAML-serialised structure, aligning with the Linux Foundation Bitol initiative.
- Defines reusable sections for:
 - Contract metadata (identifier, type, lifecycle stage, participants)
 - Roles & responsibilities (producer, consumer, aggregator, steward)
 - Access & licensing policies
 - Quality & service-level rules (refresh rate, retention, completeness)
 - Fully compatible with DCAT / DPROD semantics for integrated governance.

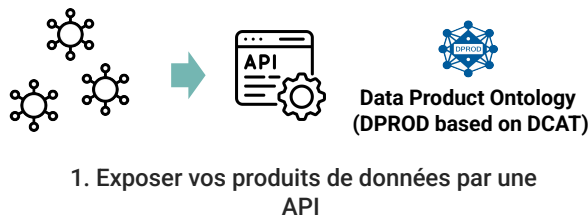
- ODCS complements DPROD by providing the governance layer of each Data Product — defining how, under which conditions, and by whom data is exposed and maintained.
- It transforms traditional textual policies into verifiable, FAIR-aligned contracts, ensuring both human and machine interpretability.

Produits et contrats de données



Que faut-il produire pour rejoindre le LUMEN Data Mesh?

We ask communities to expose – not reinvent – what they already manage, using shared standards.

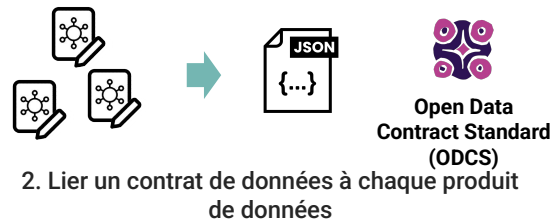


Nous n'avons pas besoin de vos données. Nous avons besoin d'une description claire et consultable de vos collections de ressources existantes.

Pour ce faire :

Publiez un point de terminaison API.
Fournissez des métadonnées conformes au modèle DPROD (basé sur DCAT) :
Dataset, Distribution, DataService
identifier, conformsTo, hasPolicy, accessURL, etc.

Cela rendra votre produit de données consultable et utilisable dans toute la fédération (par exemple dans Metasearch, WLP, etc.).



Nous n'avons pas besoin d'un avocat. Nous avons besoin de règles claires et lisibles par machine pour votre produit de données.

Pour ce faire :

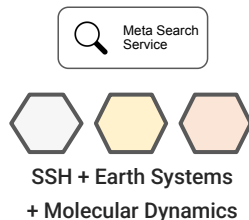
Créez un fichier YAML par produit de données en utilisant le format ODCS
Décrivez :

- Qui en est le propriétaire, le gestionnaire et l'utilisateur (propriétaire, équipe, rôles)
- Sous quelle licence, quelles règles d'accès, quelles conditions d'utilisation (hasPolicy, license)
- Comment il est maintenu : fréquence de mise à jour, période de conservation, fin de vie (slaProperties)
- Attentes en matière de qualité et de disponibilité des données (qualité, disponibilité, SLA)

Cela rend votre produit de données fiable et réutilisable dans toute la fédération (à la fois par les humains et les systèmes).

Cas d'usage du Data Mesh #1: Le Metasearch en action

Suivi des microplastiques à travers différents domaines : des molécules → transport océanique → impacts sur la santé et les politiques

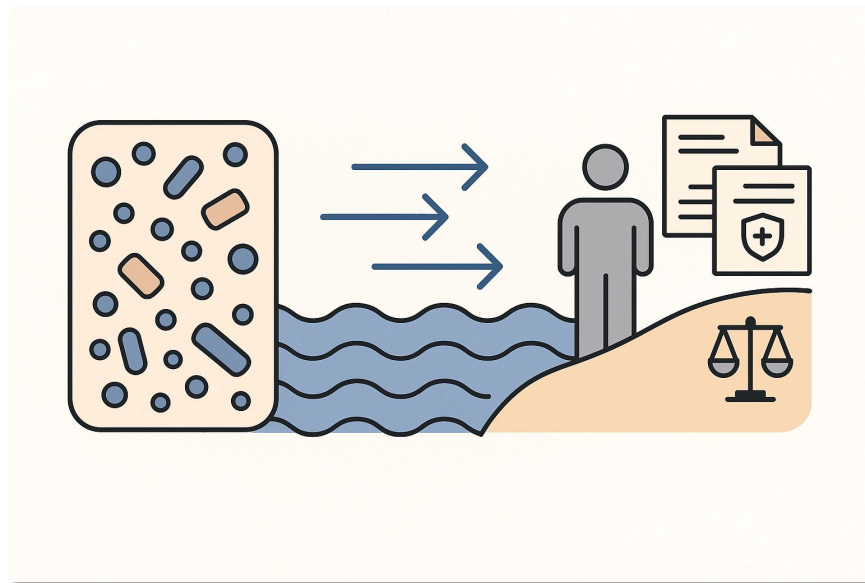


En tant que chercheur en santé environnementale, j'étudie l'impact des microplastiques sur les populations humaines dans les zones côtières.

Grâce à LUMEN, j'identifie un produit de données FAIR sur les simulations d'interactions moléculaires modélisant la dégradation des composés plastiques dans l'eau de mer.

Je le relie ensuite à un modèle de transport maritime basé sur le climat provenant de Earth System et enfin, je trace les mentions de ces composés dans les documents politiques et les études de santé publique exposés par la communauté SSH.

Cette combinaison m'aide à construire un modèle d'évaluation des risques transdisciplinaire couvrant les molécules, les mouvements et la signification.



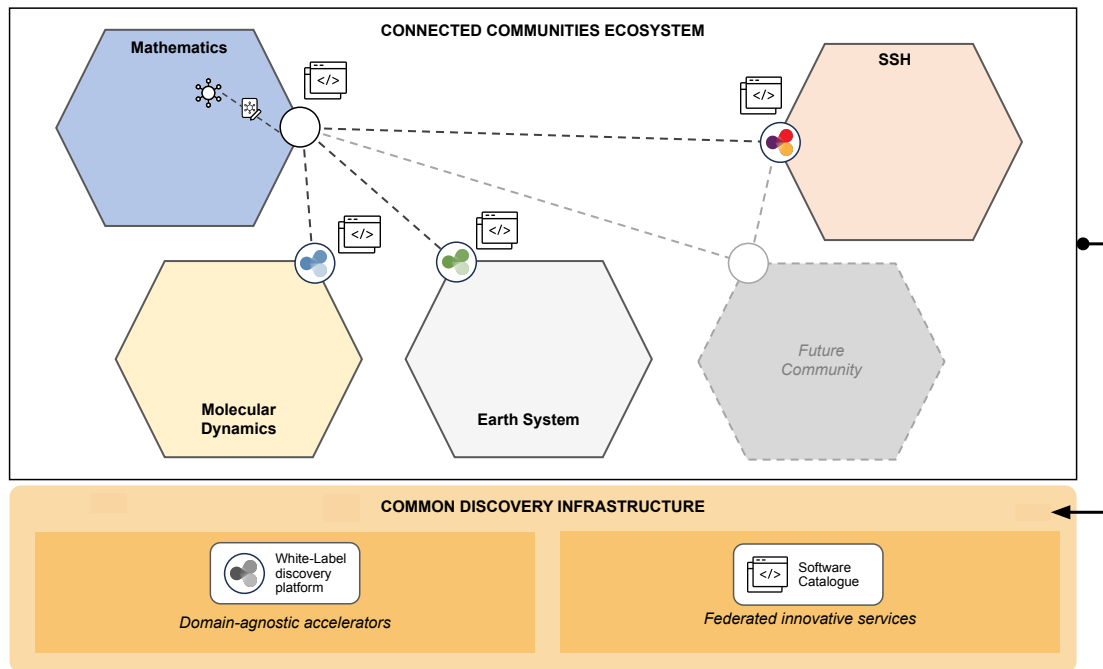
Cas d'usage du Data Mesh #2: WLP + Software catalog

Un produit de données partagé via le maillage de données LUMEN

Dans LUMEN, le catalogue de logiciels développé dans le WP6 sert de source de données canonique pour les métadonnées des logiciels de recherche. Au lieu de créer un système de collecte parallèle au sein de la plateforme en marque blanche (WP4), le catalogue est présenté comme un produit de données au sein du maillage de données.

Cette intégration garantit que les descriptions des logiciels, les identifiants et les métadonnées FAIR sont automatiquement réutilisés sur toutes les plateformes de découverte (GoTriple et autres instances en marque blanche). Elle illustre une véritable fédération entre les WP : un service gère le contenu, un autre le met à disposition, les deux étant reliés par un contrat commun et réglementé.

Cette approche évite les doublons, garantit l'alignement des schémas et des API et renforce la durabilité de l'écosystème.



Cas d'usage du Data Mesh #3: EOSC Synergies

Alignement du maillage de données LUMEN avec le catalogue de ressources du nœud EOSC de l'UE

Nous avons identifié une opportunité claire d'alignement entre le catalogue de ressources du nœud européen EOSC et le cadre LUMEN Data Mesh.

Tous deux visent à rendre les produits de recherche (données, publications et logiciels) consultables et accessibles dans tous les catalogues fédérés.

Dans LUMEN, cette exposition est structurée à travers deux mécanismes complémentaires :

- Les produits de données, qui fournissent des métadonnées, des points d'accès et un alignement FAIR ;
- Les contrats de données, qui définissent les conditions de publication et la compatibilité avec les machines.

Ensemble, ils offrent un modèle pratique permettant aux communautés d'enregistrer et de mettre à jour leurs produits de recherche au sein de la fédération EOSC, garantissant ainsi une visibilité « au-delà du nœud », en parfaite adéquation avec la vision du catalogue fédéré.

Registration of EOSC Research Product Catalogues in the EOSC EU Node

Version 3.0, September 5, 2025

Abstract The EOSC EU Node (EEN) Resource Catalogue is designed to offer discovery and access to the map of resources made available in the EOSC Federation via the EOSC Nodes, including the EEN itself. In particular, it aggregates research product profiles (publications, research data, research software) collected from the research product catalogues of EOSC Nodes. The resulting Catalogue is accessible via the EOSC Resource Hub discovery portal (and open APIs), and can be used to discover, access, and monitor EOSC research products across the EOSC Nodes in the Federation.

This document outlines the EOSC Resource Catalogue interoperability guidelines that EOSC Node Research Product Catalogue providers should follow to register their catalogues with the EEN Resource Catalogue, making their research products discoverable, visible, and accessible beyond their Node's boundaries.

Registration of EOSC Research Product Catalogues in the EOSC EU Node,
September 2025

<https://zenodo.org/records/15516020>

3 pilotes pour tester et affiner les développements

Toute la communauté scientifique est la bienvenue pour participer!

Octobre 2026 – avril 2027 : Cas d'utilisation 1 - Qualité de l'expérience utilisateur

Cet UC adaptera la plateforme GoTriple à de nouvelles communautés d'utilisateurs dans les domaines des sciences du système terrestre et de la dynamique moléculaire, en testant son applicabilité au-delà des sciences sociales et humaines (SSH). Il introduira des concepts de marque blanche, évaluera les fonctionnalités de découverte essentielles et intégrera des outils spécifiques au domaine afin d'améliorer l'expérience utilisateur dans toutes les disciplines. L'objectif consiste à obtenir rapidement des informations sur l'amélioration de l'expérience utilisateur et la flexibilité de la plateforme.

Octobre 2026 - avril 2027 : Cas d'utilisation 2 - Outils avancés pour la production de ressources

Axé sur le soutien à la création et à la conservation de contenu scientifique, ce cas d'utilisation développera et testera des outils pour la rédaction collaborative, l'édition et l'amélioration des liens entre les publications, les ensembles de données et les logiciels. Il abordera les questions de qualité, de transparence et de reproductibilité dans les résultats de recherche.

3 pilotes pour tester et affiner les développements

Toute la communauté scientifique est la bienvenue pour participer!

Janvier 2027 – juillet 2027 : Cas d'utilisation n° 3 – Amélioration des expériences de découverte

Ce cas d'utilisation testera des outils de découverte de pointe, tels que des assistants alimentés par l'IA, des modèles linguistiques à grande échelle et des visualisations novatrices, afin d'améliorer la manière dont les chercheurs recherchent et exploitent les connaissances scientifiques. Il favorisera la recherche interdisciplinaire et un accès plus large à la société grâce à des innovations telles que les chatbots et la recherche sémantique. Le cas d'utilisation 3 bénéficiera également des méthodes d'évaluation affinées et des pratiques d'engagement des utilisateurs mises en place lors des phases précédentes.

Merci beaucoup !



<https://lumenproject.eu>



[linkedin/company/lumen-eu](https://www.linkedin.com/company/lumen-eu)



[@lumen-eu.bsky.social](https://bsky.app/profile/@lumen-eu.bsky.social)



Subscribe to our newsletter!



**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.

