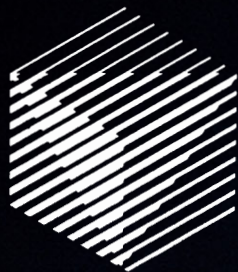


# ERCIM



# NEWS

*Special theme:*

# Advancing Open Science

Federated Infrastructures  
and Trustworthy  
Ecosystems

# Editorial Information

ERCIM News is the magazine of ERCIM. Published quarterly, it reports on joint actions of the ERCIM partners, and aims to reflect the contribution made by ERCIM to the European Community in Information Technology and Applied Mathematics. Through short articles and news items, it provides a forum for the exchange of information between the institutes and also with the wider scientific community. This issue has a circulation of about 2,000 printed copies and is also available online at <https://ercim-news.ercim.eu>.

ERCIM News is published by ERCIM EEIG  
BP 93, F-06902 Sophia Antipolis Cedex, France  
+33 4 9238 5010, [contact@ercim.eu](mailto:contact@ercim.eu)  
Director: Dominique Hazaël-Massieux,  
ISSN 0926-4981

## Contributions

Contributions should be submitted to the local editor of your country

## Copyright notice

All authors, as identified in each article, retain copyright of their work. ERCIM News is licensed under a Creative Commons Attribution 4.0 International License (CC-BY).

## Advertising

For current advertising rates and conditions, see <https://ercim-news.ercim.eu/> or contact [peter.kunz@ercim.eu](mailto:peter.kunz@ercim.eu)

## ERCIM News online edition

<https://ercim-news.ercim.eu>

## Subscription

Subscribe to ERCIM News by sending an email to [en-subscriptions@ercim.eu](mailto:en-subscriptions@ercim.eu)

## Editorial Board:

Central editor: Peter Kunz, ERCIM office ([peter.kunz@ercim.eu](mailto:peter.kunz@ercim.eu))

## Special Theme guest editors:

- Leonardo Candela (CNR-ISTI)
- Roberto Di Cosmo (Inria and University Paris Cité)

## Local Editors:

- Ferran Argelaguet, Inria, France ([ferran.argelaguet@inria.fr](mailto:ferran.argelaguet@inria.fr))
- Andras Benczur, SZTAKI, Hungary ([benczur@info.ilab.sztaki.hu](mailto:benczur@info.ilab.sztaki.hu))
- José Borbinha, Univ. of Technology Lisboa, Portugal ([jlb@ist.utl.pt](mailto:jlb@ist.utl.pt))
- Monica Divitini, NTNU, Norway ([divitini@ntnu.no](mailto:divitini@ntnu.no))
- Marie-Claire Fogue, ERCIM/W3C ([mcf@w3.org](mailto:mcf@w3.org))
- Lida Harami, ICS-FORTH, Greece ([lida@ics.forth.gr](mailto:lida@ics.forth.gr))
- Athanasios Kalogeras, ISI, Greece ([kalogeras@isi.gr](mailto:kalogeras@isi.gr))
- Georgia Kapitsaki, Univ. of Cyprus, Cyprus ([gkapi@cs.ucy.ac.cy](mailto:gkapi@cs.ucy.ac.cy))
- Annette Kik, CWI, The Netherlands ([Annette.Kik@cwi.nl](mailto:Annette.Kik@cwi.nl))
- Alexander Nouak, Fraunhofer-Gesellschaft, Germany ([alexander.nouak@iuk.fraunhofer.de](mailto:alexander.nouak@iuk.fraunhofer.de))
- Laura Panizo, University of Malaga ([laurapanizo@uma.es](mailto:laurapanizo@uma.es))
- Erwin Schoitsch, AIT, Austria ([erwin.schoitsch@ait.ac.at](mailto:erwin.schoitsch@ait.ac.at))
- Thomas Tamisier, LIST, Luxembourg ([thomas.tamisier@list.lu](mailto:thomas.tamisier@list.lu))
- Maurice ter Beek, CNR-ISTI, Italy ([maurice.terbeek@isti.cnr.it](mailto:maurice.terbeek@isti.cnr.it))

# Contents

## JOINT ERCIM ACTIONS

- 4 Message from the President
- 5 ERCIM “Alain Bensoussan” Fellowship Programme

## SPECIAL THEME

Introduction to the Special Theme:

- 6 Advancing Open Science: Federated Infrastructures and Trustworthy Ecosystems  
by the guest editors Leonardo Candela (CNR-ISTI) and Roberto Di Cosmo (Inria and University Paris Cité)

## Research assessment and scholarly representation

- 8 The OpenAIRE Graph: Enabling Open Research Intelligence Using Open Data  
by Paolo Manghi (CNR-ISTI)
- 9 Rethinking Researcher Profiles in the Research Assessment Transition Era: The OpenAIRE Approach  
by Stefania Amodeo and Zenia Xenou (OpenAIRE AMKE)
- 10 BibTexViz: Visualizing Research Productivity with Open Science Data  
by Jose-Miguel Horcas (ITIS Software, Universidad de Málaga)

- 12 A 360° View of Open Science in Europe – Indicators, Narratives, and AI-enhanced Policy Intelligence  
by Tereza Szybisty (OpenAIRE Amke)

From FAIR data to AI-ready workflows

- 14 From Data to Knowledge: Open Pipelines for Collaborative Production Research  
by Leon Gorissen and Philipp Brauner (RWTH Aachen University) Germany)

- 16 Scalable Platforms for Reusing Research Infrastructure Data in Open Science  
by Magdalena Brus (EGI Foundation), Ville Tenhunen (EGI Foundation), Gergely Sipos (EGI Foundation), on behalf of the RI-SCALE consortium

- 17 MOEBA-BIO: An Open and Extensible Framework for Evolutionary Biclustering in Biomedicine**  
by Adrián Segura Ortiz, José García-Nieto (ITIS Software, University of Málaga) and Laetitia Jourdan (CRISTAL Laboratory, University of Lille)

[Semantic foundations and knowledge graph infrastructures](#)

- 19 Reinforcing Open Science in Biodiversity through Semantic Knowledge Graphs**  
by Yannis Marketakis, Eleni Tsouloucha, Athina Kritsotaki, and Yannis Tzitzikas (FORTH-ICS)

- 21 Unlocking Machine-Composability with the EOSC Interoperability Framework Registry**  
by Alessia Bardi (CNR-ISTI, OpenAIRE AMKE), Konstantina Galouni (ARC) and Paolo Manghi (CNR-ISTI, OpenAIRE AMKE)

[Governance, skills, sovereignty, and ethical foundations](#)

- 22 Human-Centered Threat Modeling in Web Standardization at W3C**  
by Simone Onofri (W3C) and Giovanni Corti (FBK)

- 24 Co-Creation as Infrastructure for Open Science Data Spaces**  
by Rita Stampfl (University of Applied Sciences Burgenland), Silke Palkovits-Rauter (University of Applied Sciences Burgenland)

- 25 Services for Open Science Education and Skills Development: Beyond Technical Training**  
by Katharina Flicker (TU Wien, SBA), Ilire Hasani-Mavriqi (TU Graz), Dimitri Prandner (JKU)

- 27 Shaping the Czech National Infrastructure for Research Data through Community Consensus**  
by Martin Dvořák, Pavlína Špringerová and Matej Antol (Masaryk University)

- 29 Creative Commons Licences in the Age of AI: Challenges and Opportunities**  
by Daniel Spichtinger (Ludwig Boltzmann Gesellschaft/University of Vienna/independent researcher)

- 30 LICORICE: Deploying Privacy-Enhancing Technologies for Europe's Digital Sovereignty**  
by Diana E. Jimenez-Bejarano (Dialog), Stephan Krenn (AIT Austrian Institute of Technology), Antonio Kung (Dialog), and Angel Palomares Perez (Advanced Computing SL)

[Operational experiences with federated science gateways](#)

- 32 The EOSC EU Node: An Evolving Gateway to Open Science in Europe**  
by Alane Brunschweiler, Maja Dolinar and Fay Meimaraki (OpenAIRE AMKE)

- 33 The EOSC Core Innovation Sandbox: A Launchpad for Europe's Open Science Federation**  
by Federico Drago (EGI Foundation) and Nicola Fiore (EGI Foundation)

- 35 EOSC Data Commons: Building Europe's Next-Generation Research Data Infrastructure**  
by Enol Fernández, Ilaria Fava and Xavier Salazar (EGI Foundation)

- 36 ENVRI-Hub: A Science Gateway for Open Environmental Research**  
by Federico Drago (EGI Foundation), Delphine Dobler (Euro-ARGO ERIC) and Ulrich Bundke (Forschungszentrum Jülich / IAGOS AISBL)

- 38 D4Science: An Enabling Infrastructure for Open Science**  
by Massimiliano Assante, Luca Frosini, Francesco Mangiacrapa, and Pasquale Pagano (CNR-ISTI)

- 39 Supporting Open Science in Virtual Research Environments: The DAVE Experience**  
by Andrea Dell'Amico, Alfredo Oliviero, Giancarlo Panichi, Biagio Peccerillo (CNR-ISTI) and Marco Procaccini (CNR-IGG)

**ANNOUNCEMENTS**

- 41 HORIZON Europe: ERCIM Project Support Services**
- 41 Summer School on Informatics Education Research**
- 42 Dagstuhl Seminars and Perspectives Workshops**
- 42 INESC TEC International Visiting Researcher Programme 2026**
- 43 EUCNC 6G Summit**

**NEXT ISSUE**

**ERCIM News 145, July 2026**  
Special theme:  
**E-values** — Statistical Testing for the 21st Century

## Message from the President

Since its foundation in 1989, ERCIM has been a dynamic organisation in computer science and mathematics. As a networked organisation of excellent institutes, many activities have taken place and important results have been achieved. These continue today and will do so in the years ahead. I am grateful and proud to have been elected as President for a two-year term, and I look forward to working with the newly chosen Board and all ERCIM members.

What are the key focuses for ERCIM in the coming period? While we already have many activities, we will place additional emphasis on being an active network that jointly carries out research, engages with European policy and institute management, and addresses important issues like ethics and the autonomy of the EU.

Many temporary, ad hoc networks exist in our fields of science nowadays, often driven by short-term EU project funding or by the hypes of scientific topics and breakthroughs. ERCIM consists of long-standing, excellent institutes with experience across research, hypes, innovation, and longer-term goals. In this way, we distinguish ourselves as a solid European organisation in two of the most influential areas of science in current society: computer science and mathematics. Building on this, increasing the joint collaboration on research and innovation will make ERCIM stand out even more than before. Not only through our Working Groups, but also through renewed initiatives for joint research projects.

As a mid-sized network, ERCIM has the advantage that members know each other and can move forward together in a range of activities. The new Board will build on this focus on intensive collaboration in the coming years. ERCIM is supported by an excellent office, which helps to manage projects and acquire funding. Through the office, we also maintain a strong link with the W3C organisation. And of course, new members are welcome to join ERCIM.

Given our long-standing position, ERCIM has a solid role in Europe, and



**Han La Poutré, CWI, was elected President of ERCIM AISBL for a term of two years at the ERCIM General Assembly held at Inria Rennes in October 2025.**

Photo: Ivar Pel

we aim to strengthen connections with the European Commission. One of our more recent initiatives are the Visionary Events, where important developments in science and technology are highlighted and the future challenges, impacts, threats, and opportunities are addressed. In this way, we aim to connect to the European Commission, the managements of institutes, as well as to other decision makers and, of course, researchers. We have already organised events on Generative AI and Large Language Models as well as Quantum Computing, and expect to organise on Software Science in the near future.

In addition, ERCIM contributes through scientific reports and Working Groups that have their own impact in their areas, like on AI, ethics, and security and trust. These themes are not only important for the individual member institutes of ERCIM, but also for the digital autonomy of the EU.

In a time of rapid technological progress as well as geopolitical disruptions, strengthening Europe's autonomy becomes crucial. This concerns not only technology, but also economic, security, and ethical dimensions. ERCIM already contributes in these areas, and aims to strengthen this role further as a focused network of excellent institutes.

Last but not least, we would like to strengthen our connection with individ-

ual researchers. We already have our highly appreciated ERCIM News magazine, with themed issues giving quick overviews of the state of the art of important themes, by established researchers in the field. We also have an active Human Capital group, enabling ERCIM Fellowships, the Cor Baaijen thesis awards, and various diversity and inclusion activities. To reach researchers more directly, we have introduced a new ERCIM Internal Newsletter, available by email subscription. We also plan to resume broader (Fall) meetings, including a wide range of activities and Working Groups meetings.

In parallel, we are strengthening our connections with other European organisations, including ACM Europe and Informatics Europe: ERCIM is expected to participate in an upcoming, new community event of ACM Europe, and ERCIM and Informatics Europe have initiated a joint Working Group on Software Science. We also aim to expand collaboration with the Japan Science and Technology Agency (JST).

In the coming years, these activities will shape the ERCIM agenda, and the Board is committed to advancing them. I would like to thank all ERCIM members for their confidence in the new Board: Fabio Martinelli (CNR, Vice-President), Christos Koulamas (ISI, Treasurer), Alexander Nouak (Fraunhofer Gesellschaft, Secretary), Dimitris Plexousakis (ICS-FORTH, responsible for science), Monica Divitini (NTNU, responsible for human capital), Andreas Rauber (SBA, responsible for outreach), and myself. Also, I would like to thank Björn Levin (RISE) for serving ERCIM as the former President, and Gabriel David for acting as the Secretary of the ERCIM board. We owe them gratitude for their dedication to ERCIM.

Finally, I am pleased to note that I am the fourth ERCIM President via which CWI contributes to ERCIM's leadership, after Cor Baayen, Gerard van Oortmerssen, and Jos Baeten. I look forward to our joint activities in the coming years and hope to meet many of you on various occasions.

*Han La Poutré*

# ERCIM “Alain Bensoussan” Fellowship Programme

*The ERCIM Postdoctoral Fellowship Programme is one of the flagship initiatives of ERCIM. Open to young researchers from around the world, the programme covers a broad range of fields in computer science and applied mathematics.*

The fellowship scheme aims to help young scientists deepen their knowledge of European research structures and networks, while gaining valuable experience within leading European research institutions. Fellowships have a duration of 12 months, with the possibility of extension, and are hosted by one of the ERCIM member institutes.

## Hosting Institutions

Only ERCIM members can host fellows. When an ERCIM member is a consortium, the hosting institute may be any of its member organisations. When an ERCIM member is a funding body, the hosting institute may be one of its affiliated institutions.

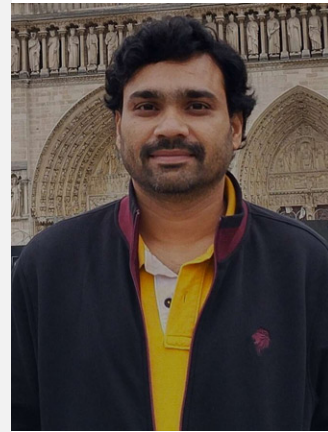
Fellowships are offered according to the needs and available funding of the member institutes. Fellows are appointed either through a stipend (a research training agreement) or a work contract, depending on the hosting institute. The type of contract and the

“

The ERCIM “Alain Bensoussan” Fellowship offered a remarkable opportunity to collaborate with leading international researchers, gain cross-cultural experience, and engage with state-of-the-art advancements in embedded AI and hardware acceleration. A particularly rewarding aspect was the Research Exchange Programme, which enabled visits to other ERCIM institutes—fostering deeper collaboration and broadening research perspectives.



Ramesh SAMBANGI  
Former ERCIM Fellow



monthly allowance or salary vary by host institution.

ERCIM encourages applications from researchers both in academia and in industry.

## Why apply for an ERCIM Fellowship?

The Fellowship Programme enables talented early-career scientists from all over the world to work on challenging research problems alongside leading European experts. In addition to research excellence, the programme fosters collaboration and knowledge exchange within the European research community.

The programme offers ERCIM Fellows the opportunity to:

- Work with internationally recognised experts;
- Gain a deeper understanding of European research structures and networks;
- Become familiar with the working conditions in leading European research centres;

- Promote cross-fertilisation and cooperation between research groups working in similar areas across Europe.

## Equal Opportunities

ERCIM is committed to ensuring equal opportunities and promoting diversity. Candidates are not discriminated against on the basis of race, colour, religion, gender, national origin, age, marital status, or disability.

## Conditions

Candidates must:

- Have obtained a PhD degree within the last eight years (prior to the application deadline), or be in the final year of their doctoral studies with an outstanding academic record. Proof of the PhD degree must be provided before the start of the fellowship;
- Be fluent in English.

## Application deadlines

Applications are accepted twice a year, with deadlines on 30 April and 30 September.

Since its inception in 1991, more than 800 fellows have participated in the programme.

The Fellowship Programme is named in honour of Alain Bensoussan, former president of Inria, one of the three founding institutes of ERCIM.

<http://fellowship.ercim.eu>



“

The ERCIM Fellowship has been an outstanding opportunity to broaden my research horizons, collaborate with leading European institutes, and gain hands-on experience in cutting-edge hardware security. The program not only enhanced my technical expertise but also expanded my professional network, opening doors to long-term collaborations and opportunities. I strongly recommend it to researchers who want to combine scientific excellence with international exposure.



HENDARMAWAN  
Former ERCIM Fellow



Introduction to the Special Theme

## Advancing Open Science: Federated Infrastructures and Trustworthy Ecosystems

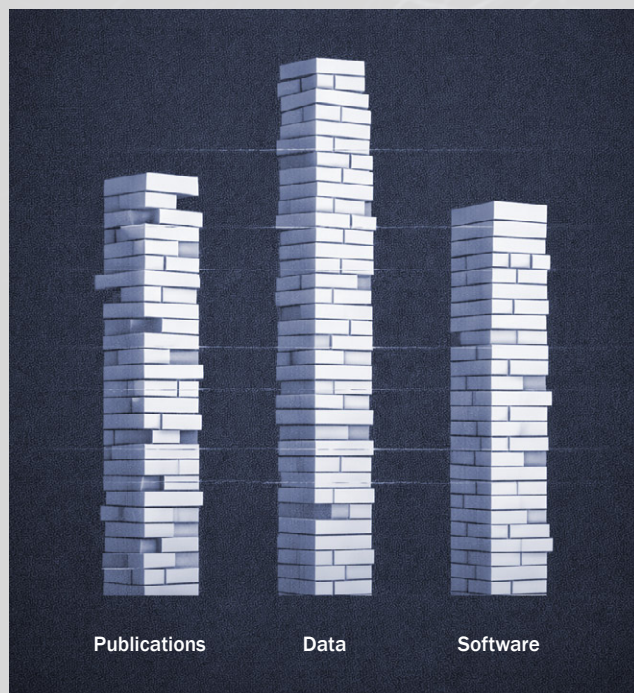
by the guest editors Leonardo Candela (CNR-ISTI) and Roberto Di Cosmo (Inria and University Paris Cité)

Open Science is a broad and evolving movement. The UNESCO framework describes it as an inclusive approach aimed at making scientific knowledge openly available, accessible, and reusable for everyone, opening the processes of knowledge creation and evaluation to stakeholders beyond the traditional research community [1]. Today, Open Science is no longer merely a normative ideal: it has become an operational requirement embedded in national strategies, funding conditions, and research assessment reforms across Europe and beyond.

Yet the very breadth of Open Science is also its greatest challenge. The movement rests on several distinct pillars, each with its own history, infrastructure landscape, and degree of maturity — and each exposed to the same structural risk: fragmentation.

### The three pillars — and the fragmentation trap

The oldest pillar is **open access to publications**. Decades of effort have produced undeniable progress, but also a cautionary tale. Because coordination came late, the landscape is now highly fragmented: OpenDOAR counts over 6,000 open access repositories worldwide, each requiring its own infrastructure, archival, backup, and metadata curation. Content is duplicated, metadata is inconsistent, and the cost of maintaining this



Open Science is built on three foundational pillars, but each evolved with different levels of fragmentation and coherence.

patchwork is borne many times over. The recent move to fund, via national grants, the EU-originated Open Research Europe journal illustrates how difficult it is to retrofit coherence onto an ecosystem that grew without a shared architectural plan.

The second pillar, **open research data**, has benefited from the lessons of publications and from the early adoption of the FAIR principles. Yet a similar proliferation of platforms and curation challenges is already visible, with a very long tail of research data that struggles to find a sustainable home. National initiatives such as Recherche Data Gouv in France and the PLATICA project in Spain point toward a promising model: shared, mutualized infrastructures that host curated research data as a public good, rather than leaving each institution to build and maintain its own silo.

The third pillar — **research software** — has long pre-existed the others, since software has been at the heart of scientific computation for decades. Yet it was recognized as a pillar of Open Science only very recently. The French Second National Plan for Open Science (2021) was the first national strategy to dedicate a full chapter to software, establishing measures for archiving, referencing, and citing source code, creating a national research software award, and providing explicit support for Software Heritage as a key infrastructure [2]. Spain is now actively building on this momentum, as evidenced by the discussions at the recent second national days on Open Science held in Aranjuez in March 2026.

For software, there is a unique opportunity to avoid the fragmentation that has plagued publications and data. Software Heritage was designed from the outset as a universal, open, non-profit archive for all software source code. It already preserves over 28 billion source files from more than 430 million projects collected across over 5,000 code hosting and distribution platforms worldwide, assigning intrinsic, cryptographically strong identifiers (SWHIDs, now standardized as ISO/IEC 18670). This provides a single, shared layer for archiving, referencing, describing, and citing software — a foundation that Open Science policy can build on directly, without the need to reconcile thousands of independent local repositories after the fact.

### Federating from the top: promise and friction

Alongside bottom-up infrastructure efforts, Europe has invested heavily in top-down coordination through the European Open Science Cloud (EOSC), which aims to federate existing services into an interoperable, cross-border research environment. Several contributions to this issue illustrate both the promise and the complexity of this endeavour.

Yet federation by decree is hard. Even in countries with active EOSC engagement, surveys show that a majority of researchers still store data primarily on personal computers, and awareness of federated infrastructure remains low. The gap between policy ambition and daily research practice is real, and bridging it requires not just technical platforms but sustained investment in skills, incentives, and institutional culture change.

#### A map of the current landscape

The contributions collected in this special theme offer a cross-section of the current European effort, organised into five thematic clusters.

A first cluster addresses **research assessment and scholarly representation**. The OpenAIRE Graph (Manghi) provides a community-governed scholarly knowledge graph treating datasets and software as first-class outputs, offering an open alternative to proprietary research intelligence. MyResearchFolio (Amodeo and Xenou) builds on this to support richer researcher profiles aligned with responsible assessment principles, while BibTexViz (Horcas) demonstrates visual analytics for open bibliographic data. The EOSC Open Science Observatory (Szybisty) combines indicators, national narratives, and AI-assisted analysis to monitor Open Science progress across Europe.

A second cluster explores the transition **from FAIR data to AI-ready workflows**. Contributions show how shared industrial datasets can feed collaborative knowledge pipelines (Gorissen and Brauner), how compute-to-data architectures enable scalable analysis on research infrastructures (Brus et al.), and how modular, open-source research software frameworks can support advanced biomedical analytics (Segura-Ortiz et al.).

A third cluster highlights **semantic foundations and knowledge graph infrastructures** as critical enablers of interoperability, through the transformation of legacy databases into FAIR-by-design knowledge graphs (Marketakis et al.) and the evolution of the EOSC Interoperability Framework toward machine-actionable, composable service templates (Bardi et al.).

A fourth cluster addresses the **governance, skills, sovereignty, and ethical foundations** without which technical infrastructure cannot function. Contributions cover human-centred threat modelling (Onofri and Corti), structured co-creation in data spaces (Stampfl and Palkovits-Rauter), Open Science education beyond purely technical skills (Flicker et al.), the Czech national experience with FAIR adoption (Dvořák et al.), the tension between Creative Commons licences and AI training (Spichtinger), and privacy-enhancing technologies for secure cross-border data sharing (Jimenez-Bejarano et al.).

The fifth and final cluster presents **operational experiences with federated science gateways**, including the EOSC EU Node (Brunschrweiger et al.), the Innovation Sandbox (Drago

and Fiore), the Data Commons (Fernández and Fava), the ENVRI-Hub for environmental research (Drago et al.), the D4Science virtual research environments (Assante et al.), and the DAVE conversational AI assistant for navigating complex research workflows (Dell'Amico et al.).

#### Looking ahead

Taken together, these contributions make clear that the next phase of Open Science will be defined not just by openness, but by trustworthiness and integration. Several priorities stand out.

First, avoiding fragmentation must become a conscious design principle, not an afterthought. For each pillar of Open Science — publications, data, and software — the question is whether we build shared, mutualized infrastructure from the start or spend decades trying to harmonize a patchwork.

Second, research assessment must formally recognize the full range of research outputs — datasets, software, workflows — alongside publications, moving away from proprietary metrics toward transparent, community-governed research intelligence.

Third, the intersection of open licensing and AI training remains legally ambiguous. As AI models increasingly consume open research data and code, robust opt-in/opt-out mechanisms and legal clarity are urgently needed.

Finally, long-term financial sustainability for community-governed infrastructure remains an open problem. Short-term project funding cannot secure the digital commons on which European research increasingly depends.

If the first phase of Open Science was about making research outputs accessible, the present phase is about making research ecosystems interoperable, intelligent, and trustworthy. The contributions in this issue offer both concrete experiences and forward-looking perspectives on how Europe is working to make that vision a reality.

#### References:

- [1] UNESCO (2022) An introduction to the UNESCO Recommendation on Open Science. <https://doi.org/10.54677/XOIR1696>
- [2] Second French Plan for Open Science, Generalising Open Science in France, 2021-2024 <https://www.ouvri.lascience.fr/second-national-plan-for-open-science/>

#### Please contact:

Leonardo Candela, CNR-ISTI, Italy  
[leonardo.candela@isti.cnr.it](mailto:leonardo.candela@isti.cnr.it)

Roberto Di Cosmo  
Inria and University Paris Cité, France  
[roberto@dicosmo.org](mailto:roberto@dicosmo.org)

# The OpenAIRE Graph: Enabling Open Research Intelligence Using Open Data

by Paolo Manghi (CNR-ISTI)

**The OpenAIRE Graph provides an open, community-governed data infrastructure for research intelligence, enabling transparent and auditable use of scholarly data beyond proprietary systems.**

The global research ecosystem is calling for a structural transition from proprietary, opaque systems for research intelligence to open, community-governed infrastructures. At the centre of this shift is the need to reclaim how scholarly data is collected, connected, and used to inform research evaluation and policy. The OpenAIRE Graph addresses this challenge by providing a large-scale, openly accessible scholarly knowledge graph that treats publications, data, and software as first-class research outputs. As a community-governed infrastructure, it establishes a transparent and auditable foundation for Open Research Intelligence.



## The Lock-in of Proprietary Research Intelligence

The global research community is currently trapped in a costly and paradoxical model where public funds fuel scientific discovery while simultaneously sustaining the multi-billion-euro “drain” of commercial vendors who control a large portion of the scholarly record. Access to scientific articles, books, and critical bibliographic metadata databases, such as Scopus and Web of Science, is provided through expensive subscriptions characterized by restrictive reuse constraints and a lack of transparency. Because research intelligence, institutional evaluations, and national comparisons are grounded in these proprietary databases, the scientific community is forced to rely on “black boxes” that unilaterally decide what is measured, how it is measured, and what remains visible. These platforms use data that cannot be fully audited and methods that cannot be challenged, often driven by commercial priorities rather than the public interest. It is a profound paradox that while public funds invest in research as a “public good”, scientific reviews and the critical indicators used to assess its quality and monitor its impact are provided by closed, toll-gated proprietary systems.

## The OpenAIRE Graph

In this reforming scenario, the OpenAIRE Graph [L1] has emerged and matured as a fully-fledged solution to ensure a safe and incremental transition from a closed-data to an open-data scholarly communication ecosystem. Funders, countries, and institutions are adopting it for a shift to Open Research Intelligence, researchers for research on bibliometrics and sci-

ence of science, and companies for commercial purposes. Its main features can be summarised as:

### *Global Coverage and Open Data*

The OpenAIRE Graph provides a 360-degree view of the research lifecycle, embracing the diversity raised by Open Science, hence treating research data and software as first-class citizens alongside traditional articles and scientific literature at large. The Graph consists of a collection of bibliographic metadata about research outcomes, connected by semantic relationships, including citations from publications to data and software, and data-to-data links, affiliations, participation to projects, etc. To ensure open access coverage and high-quality data, the Graph aggregates metadata and relationships from over 2,100 direct metadata sources (155,000 counting indirect data sources via aggregators), including Crossref, DataCite, ROR, ORCID, PubMed, ArXiv, and thousands of repositories, CRIS systems, and OA publishers. Aggregated metadata flows through a data wrangling workflow involving data harmonisation, AI methods inference, full-text mining, and deduplication [1]. As of the latest statistics (March 2026), the OpenAIRE Graph counts ~345Mi research products (~215Mi publications, ~95Mi research datasets, ~800K research software, 33Mi other products), 368 funders, ~450K organizations, 4 Mi grants and over 7 billion relationships. The collection is accessible via open APIs and data dumps [L3].

### *A Community-Governed Pillar of Open Research Intelligence*

Unlike commercial and non-commercial counterparts, the OpenAIRE Graph is operated by OpenAIRE AMKE [L2], a non-profit, membership-based organization established to ensure that research intelligence is sustained and steered as a public infrastructure by its members. OpenAIRE counts 53 members across 36 countries, supported by a network of experts active since 2009 whose global reach extends through collaborations in Latin America, Canada, Japan, Korea, and China. The Graph is co-designed, co-developed, and co-operated by its members, ensuring that the technical roadmap, design decisions, and sustainability are driven by community demands, ethical and political choices rather than profit motives. By treating research intelligence as infrastructure rather than a service, OpenAIRE provides a stable foundation that connects diverse research ecosystems without the risk of commercial lock-in.

### *Enabling Next-Generation Open Research Intelligence*

As of today, the OpenAIRE Graph serve open, auditable data, as input to a rich catalogue of Research Intelligence services for policymakers and researchers [L3]. Examples are services for research discovery (OpenAIRE EXPLORE and CONNECT), for monitoring Open Science and research impact (OpenAIRE MONITOR and EOSC OSObservatory, see article by T. Szybisty in this issue on page 12), and for managing Open Science CoARA-oriented research profiles (ResearchFolio, see following article by S. Amodeo).

## Conclusion

The OpenAIRE Graph leverages a shift from “intelligence as a service” (outsourced design, no transparency, no service operation) to “intelligence as infrastructure” (co-design, data transparency, service operation). Its adoption in real-case applications is a concrete example of how, by investing in a community-governed, open-data fabric and services, institutions

and ministries can move away from opaque commercial representations and ensure that the values of Open Science become a durable part of the global research landscape.

**Acknowledgments:** The OpenAIRE Graph is the result of research and operational efforts from Michele Artini, Claudio Atzori, Miriam Baglioni, Alessia Bardi, Michele De Bonis, Sandro La Bruzzo, and Andrea Mannocci from CNR-ISTI; Giambattista Bloisi, Ioanna Grypari, Natalia Manola, Harry Dimitropoulos, Yannis Foufoulas, Myrto Kallipoliti, Antonis Lempesis, Leonidas Pispiringas, and Giacomo Trombi from OpenAIRE AMKE; and Marek Horst, Michal Politowski, and Sebastian Tymkow from University of Warsaw.

**Links:**

- [L1] <https://graph.openaire.eu>
- [L2] <https://www.openaire.eu>
- [L3] <https://catalogue.openaire.eu>

**Reference:**

- [1] P. Manghi, “Challenges in building scholarly knowledge graphs for research assessment in open science.” *Quantitative Science Studies* 5, 4 (2024): 991–1021. doi:10.1162/qss\_a\_00322.

**Please contact:**

Paolo Manghi, CNR-ISTI, Italy  
paolo.manghi@isti.cnr.it

# Rethinking Researcher Profiles in the Research Assessment Transition Era: The OpenAIRE Approach

by Stefania Amodeo and Zenia Xenou (OpenAIRE AMKE)

*OpenAIRE introduces a new approach to researcher profiles, combining open data and narrative CVs to support responsible research assessment beyond traditional metrics.*

Research assessment is undergoing a fundamental transformation. For many years, academic evaluation has heavily relied on traditional publication metrics based on journal articles, citation counts, and impact factors. This approach, however, fails to capture the full range of activities that define modern research, including software development, dataset curation,

student mentorship, community engagement, and interdisciplinary collaboration. As initiatives like the Coalition for Advancing Research Assessment (CoARA) and the Barcelona Declaration on Open Research Information gain momentum, the research community is increasingly calling for more comprehensive and transparent evaluation practices that rely on open systems rather than proprietary ones.

The OpenAIRE Graph provides the open and transparent data infrastructure underpinning these developments [1].

Until recently, individual researchers remained in the background of research assessment systems, appearing only as identifiers within research output metadata. This represented a significant gap in the implementation of responsible research assessment practices.

OpenAIRE has addressed this gap by integrating researcher identities as distinct entities within the OpenAIRE Graph. This enhancement will be included in the April 2026 Graph release. Researchers with ORCID profiles become first-class entities

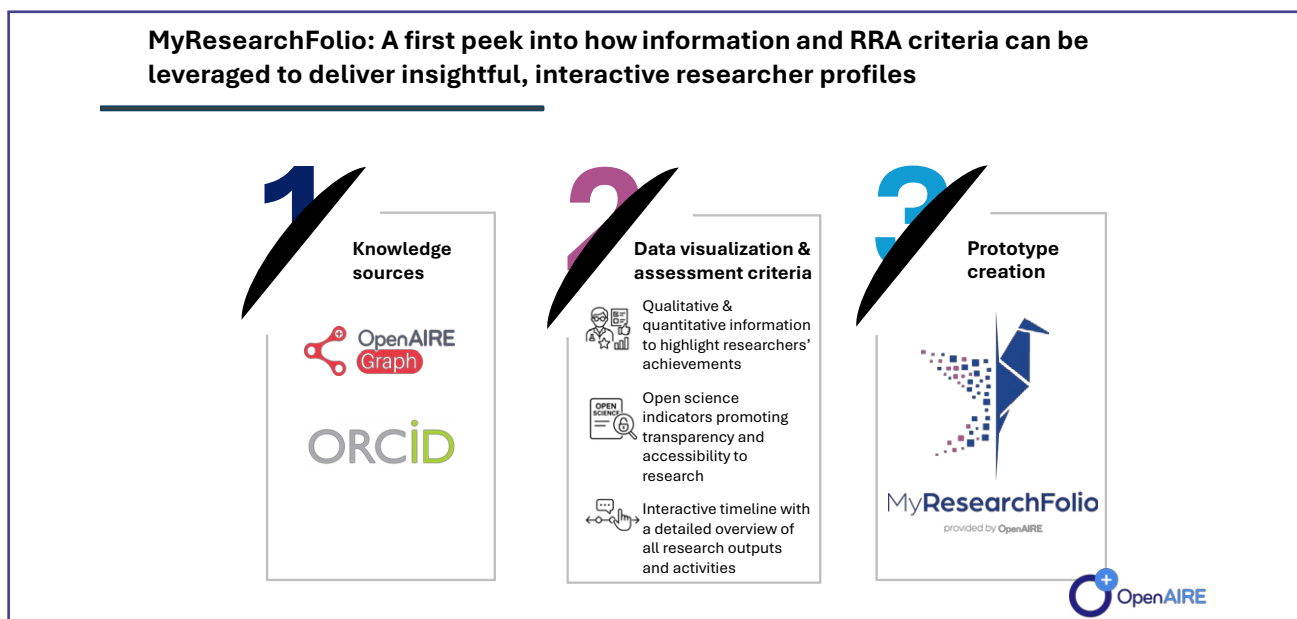


Figure 1: MyResearchFolio concept: researcher profiles are built on OpenAIRE Graph and ORCID data, combining qualitative and quantitative information with responsible research assessment (RRA) criteria, Open Science indicators, interactive visualizations, and a timeline of research outputs and activities.

in the Graph, enabling rich connections that reveal the full scope of their scholarly activities, including:

- organizational connections: employment affiliations based on ORCID profiles
- research output links: direct connections to publications, research software, datasets, and other diverse research contributions
- co-authorship networks: collaboration patterns across multiple publications
- detailed attributions: author roles, affiliations at time of publication, and corresponding author information.

The OpenAIRE Graph currently contains:

- 373 million co-author relationships connecting researchers who have collaborated on publications
- 222 million authorship links between researchers and their research outputs
- 7.5 million organizational affiliations derived from ORCID employment data
- 832,000 project participations linking researchers to funded research initiatives.

The Person entity represents a significant step toward creating a comprehensive, interconnected view of the research ecosystem. The goal is to evolve from basic researcher identification to rich, contextualized attribution that builds a complete research narrative. This means:

- Expanding the coverage of author roles to each output, including corresponding author information, the declared affiliations matched from publication metadata, and projects/funding relationships.
- Moving from static organizational connections to affiliations that capture where researchers produced specific outputs.

Building on the OpenAIRE Graph and its new Person entity foundation, MyResearchFolio [L1] is a pilot service that supports holistic and FAIR-aligned researcher profiles. The service was developed through the Horizon Europe project GraspOS [L2], and based on the results from OPUS project [L3], focused on highlighting researchers' contributions in the best light possible and on promoting Responsible Research Assessment practices.

MyResearchFolio allows researchers to create comprehensive profiles that combine quantitative indicators from the OpenAIRE Graph with qualitative narrative curriculum vitae (CVs). The concept is illustrated in Figure 1. Its key features include:

- comprehensive metadata integration: direct access to publications, datasets, software, funding information, institutional affiliations, and collaboration networks from the OpenAIRE Graph
- trusted identity management: bidirectional synchronization with ORCID, ensuring profile portability and reusability across platforms
- narrative flexibility: customizable templates that capture diverse scholarly contributions including leadership roles, interdisciplinary collaboration, mentorship activities, and community engagement
- Open Science alignment: embedded indicators measuring FAIRness, openness, reproducibility, and contributions to Sustainable Development Goals.

By integrating these elements into a single, open framework, MyResearchFolio provides a practical tool for implementing responsible research assessment principles. It enables institutions and researchers to view the research landscape through the lens of individual scholars, revealing their research outputs, career trajectories, collaborative networks, and funding history in a comprehensive and transparent manner.

These developments build on the OpenAIRE Graph infrastructure described in the article on the OpenAIRE Graph by Manghi on page 8 in this issue.

#### Links:

[L1] <https://researcherprofile.openaire.eu/>

[L2] <https://graspos.eu/>

[L3] <https://opusproject.eu/>

#### Reference:

- [1] P. Manghi, et al., "OpenAIRE Graph Dataset (10.6.0) [Data set]", OpenAIRE. <https://doi.org/10.5281/zenodo.17725827>

#### Please contact:

Stefania Amodeo

OpenAIRE AMKE, Greece

[stefania.amodeo@openaire.eu](mailto:stefania.amodeo@openaire.eu)

## BibTexViz: Visualizing Research Productivity with Open Science Data

by Jose-Miguel Horcas (ITIS Software, Universidad de Málaga)

*As the research community increasingly embraces Open Science, the need for transparent and accessible tools to understand academic productivity has become more pressing. BibTexViz [L1] bridges the gap between raw bibliographic data and insightful visual analytics, enabling researchers to explore their career trajectories while aligning with the principles of the CoARA and DORA initiatives.*

The evaluation of research productivity is undergoing a paradigm shift. Traditional assessment practices, often based on proprietary publication databases and metrics that lack transparency, are increasingly being challenged by the global movement toward Open Science. Initiatives such as the San Francisco Declaration on Research Assessment (DORA) [L2] and the Coalition for Advancing Research Assessment (COARA) [L3] advocate for more transparent, diverse, and responsible approaches to evaluating research impact. In this evolving context, enabling researchers to own, explore, and communicate their own bibliographic data has become a central concern.

Despite the abundance of digital research information, many researchers find it difficult to maintain an up-to-date and meaningful overview of their scientific output. University research portals [L4, L5] aim to address this need, yet they often

rely on proprietary databases, such as Scopus or Web of Science. These systems may exclude high-quality open-access venues or regional conferences and frequently require costly licensing agreements covered by public funds, raising concerns about reproducibility, inclusiveness, and sustainability.

### BibTexViz: From BibTeX to Insight

BibTeX remains the de facto standard for managing bibliographic references in computer science and many other disciplines. Researchers routinely curate BibTeX files for papers and grant proposals, making them an ideal foundation for analysis and visualization. BibTexViz builds on this practice by transforming standard BibTeX entries into interactive and structured visualizations that provide immediate insight into a researcher’s publication history (Figure 1).

In BibTexViz, each publication is represented as a visual element positioned along a temporal axis and enriched with semantic information such as venue type, collaboration context, or special annotations. This design enables users to generate a comprehensive visual map of their academic trajectory, revealing patterns hidden in traditional text-based lists.

### Technical Integration and Open Data

BibTexViz operates on two complementary input levels. First, it integrates directly with DBLP, the open computer science

bibliography [L6]. DBLP provides high-quality, curated bibliographic data for major journals and conferences, and has long been a cornerstone of open bibliographic infrastructure. Today, it is operated and further developed by Schloss Dagstuhl [L7], ensuring long-term sustainability and openness. By entering a researcher’s DBLP identifier, BibTexViz automatically retrieves recent and verified bibliographic data, ensuring high data quality with minimal effort from the user.

Second, recognizing that no automated database is fully comprehensive, BibTexViz allows researchers to provide custom BibTeX files. This functionality enables the inclusion of publications not indexed in DBLP, interdisciplinary outputs, or domain-specific venues. Additionally, BibTexViz supports extended BibTeX fields, allowing users to add qualitative meta-data such as tags for “Best Paper” awards, industrial collaborations, or international research stays. This customization is particularly valuable when preparing evaluation dossiers for tenure tracks, research productivity assessments, or grant applications, where contextual information complements quantitative indicators.

### Visualization Design

BibTexViz follows established visualization best practices [2], emphasizing clarity, consistency, and cognitive efficiency. Visual encodings are intentionally simple, avoiding clutter

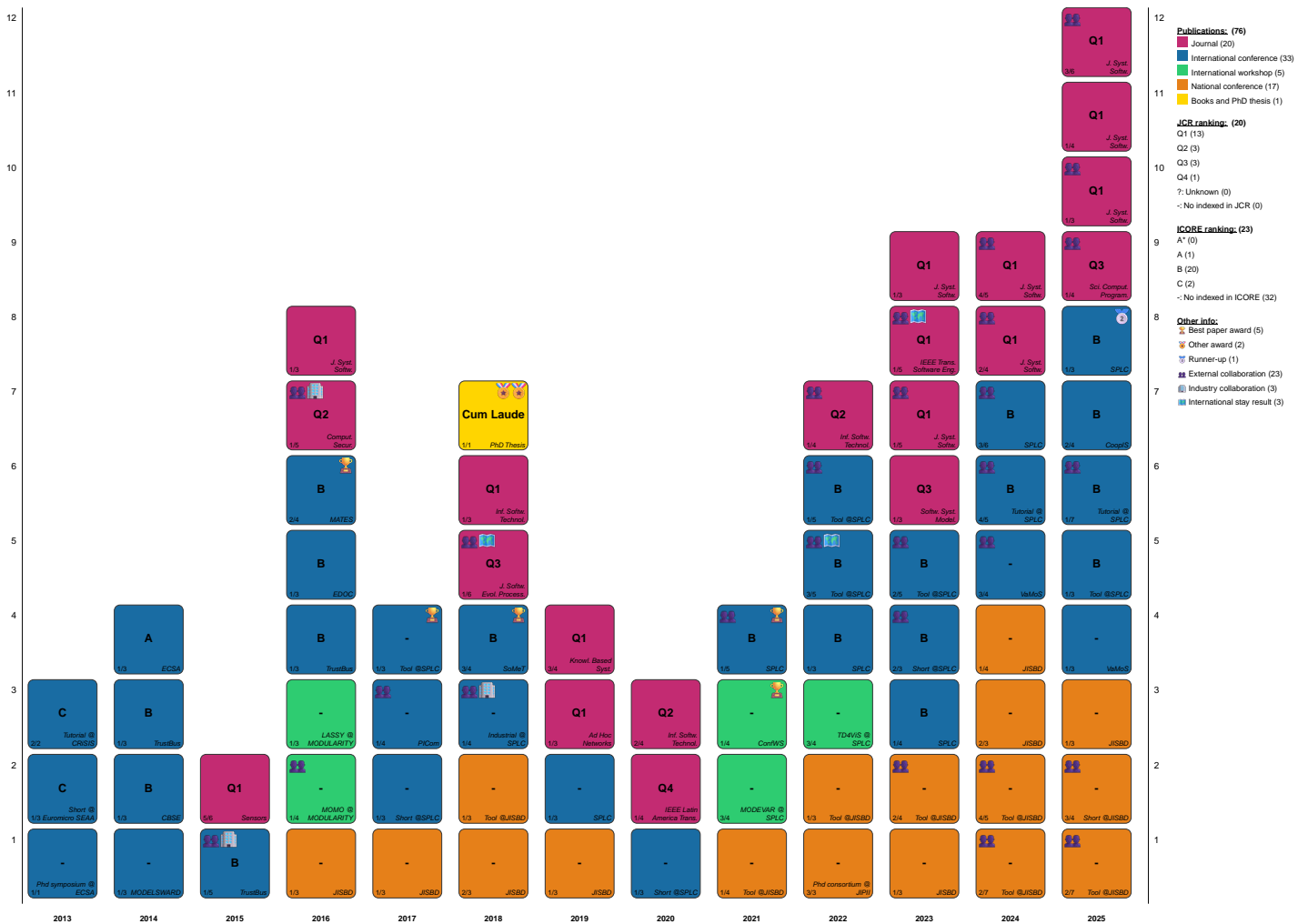


Figure 1: Interactive visualization of a researcher’s publication record generated by BibTexViz.

while conveying rich information. Color, shape, and spatial grouping are used to distinguish publication types, venues, and collaboration patterns in a way that remains interpretable even for large publication records.

Figure 1 shows an example visualization generated by BibTexViz for an individual researcher. The visualization illustrates how publications are distributed over time, how different types of venue categories are visually separated, and how milestones such as awards or major achievements can be highlighted. Such visual summaries provide an immediate overview that complements traditional publication lists in terms of interpretability and communicative power.

#### Open Science

BibTexViz is deeply rooted in the principles of Open Science and responsible research assessment. Rather than computing scores, rankings, or impact factors, the tool empowers researchers to tell the story of their research using open data and transparent visual representations. This approach aligns directly with the recommendations of DORA and COARA, which emphasize the transition from quantitative-based metrics toward contextual and qualitative evaluation practices [2].

By relying on DBLP as a primary data source and remaining fully open source, BibTexViz promotes the use of non-proprietary research infrastructures. Both the data sources and the visualization logic are transparent and inspectable, enabling reuse, verification, and integration with other tools and workflows. This openness allows BibTexViz to evolve as part of a broader ecosystem of open scholarly infrastructures rather than as a standalone, closed platform.

#### Supporting Researchers in Practice

Beyond formal evaluation contexts, BibTexViz supports everyday research activities as a tool for reflection and strategic planning [3]. Researchers can use it to analyze their publication strategies, identify under-represented venues or collaboration patterns, and communicate their work to non-expert audiences. For early-career researchers, the visualizations provide a clear view of how their research profile evolves over time, such as transitions from conference-focused outputs to more consolidated journal publications. For research groups, BibTexViz offers a bird's-eye view of collective productivity and strengths.

BibTexViz is freely available online and can be used without installation. The project is under active development, with ongoing work on additional visual encodings, enhanced customization, and support for research communities beyond computer science. Future directions include integration with other open infrastructures such as Crossref [L8], OpenAlex [L9], and open scholarly resources like Semantic Scholar [L10], further strengthening its Open Science foundations.

By combining open bibliographic infrastructures, standardized reference formats, and carefully designed visualizations, BibTexViz demonstrates how open tools can meaningfully support researchers while remaining faithful to the values of Open Science and responsible research assessment.

#### Links:

- [L1] <https://bibtexviz.github.io>
- [L2] <https://sfdora.org>
- [L3] <https://www.coara.org>
- [L4] <https://portal.delainvestigacion.uma.es>
- [L5] <https://prisma.us.es>
- [L6] <https://dblp.org>
- [L7] <https://www.dagstuhl.de/en>
- [L8] <https://www.crossref.org>
- [L9] <https://openalex.org>
- [L10] <https://www.semanticscholar.org>

#### References:

- [1] D. Hicks, et al., “The Leiden Manifesto for research metrics”, *Nature* 520, 429–431, 2015. <https://doi.org/10.1038/520429a>
- [2] T. Munzner, “Visualization Analysis and Design”, CRC Press, 2014. <https://doi.org/10.1201/b17511>
- [3] M. Zaumanis, “Research Data Visualization and Scientific Graphics: For Papers, Presentations and Proposals”, *Peer Recognized*, 2021.

#### Please contact:

Jose-Miguel Horcas  
ITIS Software, Universidad de Málaga, Spain  
[horcas@uma.es](mailto:horcas@uma.es)

## A 360° View of Open Science in Europe – Indicators, Narratives, and AI-enhanced Policy Intelligence

by Tereza Szybisty (OpenAIRE Amke)

***A new generation of Open Science monitoring in Europe connects comparable indicators with national context and AI-enhanced policy analysis to support informed and responsible decision-making.***

Open Science is now embedded in national policies, funding requirements, and research assessment debates across Europe. The challenge is no longer adoption, but understanding whether these policies are effective. Initiatives such as the Open Science Monitoring Initiative (OSMI) [L1] highlight the need for structured, transparent, and coordinated approaches to monitoring. The EOSC Open Science Observatory [L2] addresses this need by providing a framework that connects policies, practices, and impacts within a coherent and comparable model (Figure 1). The EOSC Open Science Observatory builds on the OpenAIRE Graph infrastructure described in the article by Manghi on page 8 in this issue, extending it with policy and narrative layers for monitoring and decision support.

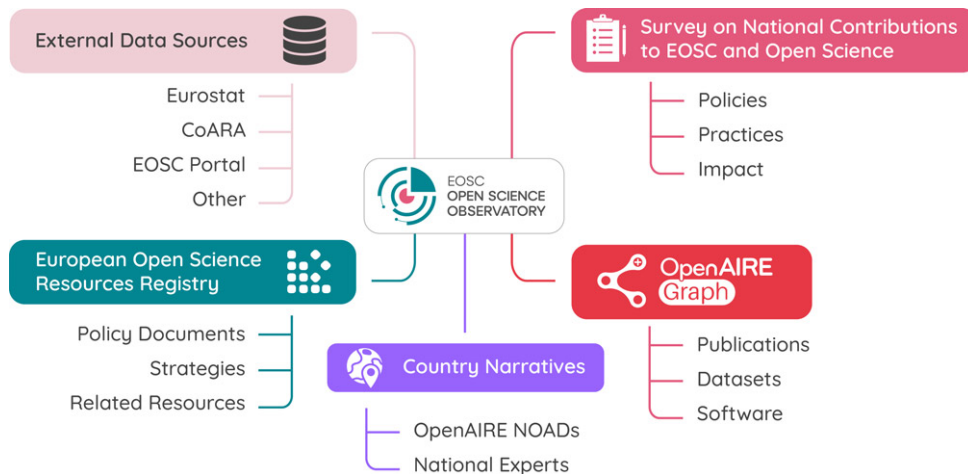


Figure 1: EOOSC Open Science Observatory Data Sources.

### The main evidence streams behind the 360° view

#### 1. Structured national monitoring through the EOOSC Steering Board survey

A key source is the annual survey on National Contributions to EOOSC and Open Science, coordinated through the EOOSC Steering Board with support from the European Commission (running annually since 2021). The survey captures comparable information across countries and is grounded in a monitoring framework (Figure 2) that organises indicators across three dimensions — policies, practices, and impacts — across eight Open Science categories: publications, data, software, services, infrastructure, skills/training, assessment, and engagement. This structured approach enables trend monitoring, and systematic identification of where enabling conditions exist (policy), where implementation occurs (practice), and what effects are visible (impact).

#### 2. Research ecosystem evidence through the OpenAIRE Graph

To complement survey-based evidence, the Observatory draws on the OpenAIRE Graph [L3] as a large-scale scholarly knowledge graph that links research outputs, organisations, projects and other entities and funding relationships. This enables analyses of Open Science practices, such as Open Access uptake or links between outputs and funding, across countries and over time. The Graph is described in detail in the article on the OpenAIRE Graph by Manghi in this issue.

#### 3. Country narratives to interpret progress and explain “why”

A crucial part of meaningful Open Science monitoring is the narrative that explains what lies behind the numbers collected through the annual survey and the OpenAIRE Graph. Indicators can show trends and levels of adoption, but they cannot explain what drove progress, which policy decisions or investments enabled change, what structural or cultural barriers persist, or what actions are planned next. To address this, the EOOSC Open Science Observatory presents dedicated country pages that integrate quantitative evidence with structured national narratives, offering a contextualised view of each country’s Open Science landscape. These

pages are directly connected to the OpenAIRE country pages maintained by the National Open Access Desks (NOADs), building on years of national reporting expertise and ensuring that monitoring data is anchored in locally grounded knowledge and coordinated European practice.

#### 4. Policy intelligence through the AI-enhanced European Open Science Resources Registry

The European Open Science Resources Registry forms a fourth key pillar of the Observatory’s 360° model. It is a curated, AI-enhanced (OpenAI) collection of national policies, strategies, and best practices supporting Open Science across Europe, offering easy filtering, rich metadata, and direct access to full texts. By integrating policy documents into the monitoring ecosystem, the Registry connects measurable trends with the formal frameworks that shape them. Its AI-supported and human-curated tools help users navigate and analyse policy texts, identify relevant provisions, and compare approaches across countries.

The Observatory also integrates selected external sources, such as Eurostat and CoARA data, further enriching the pic-

	Policies	Practices	Impacts
Publications	Countries with a National Policy	Countries with National Monitoring	Country impact Stories
Data			
Software	Countries with a Financial Strategy	Country Use Cases	Country Output Outcomes
Services			
Infrastructure	Country RPOs with a Policy	Country Investments	Correlation between Country Indicators
Skills/ Training			
Assessment	Country RFOs with a Policy	Country Outputs	Yearly Progress in Country Indicators
Engagement			

Figure 2: Monitoring Framework for National Contributions to EOOSC and Open Science.

ture and situating Open Science developments within the broader research and innovation landscape.

#### Who benefits and how the Observatory can be used

The Observatory is designed for a broad set of stakeholders, including policymakers, national Open Science coordinators, research organisations, funders, and experts. Typical use cases include:

- Tracking progress over time within a country or across Europe using the policies–practices–impacts structure.
- Comparing thematic areas (e.g., Open Access, data stewardship, research software, citizen science) to identify where enabling conditions exist but adoption lags.
- Supporting policy alignment by browsing and interpreting policy documents.
- Informing policy reform discussions, including responsible research assessment, by connecting monitoring evidence with incentive structures and implementation realities.

As Open Science monitoring evolves globally, initiatives such as the OSMI underline the need for more coordinated, transparent, and co-creation approaches. The EOSC Open Science Observatory contributes to this shift by demonstrating how monitoring can be done with and for community, combining openness, methodological transparency, validation, and FAIR data principles. By integrating indicators, narratives, policy intelligence, and trusted external evidence in an accessible and reusable way, the Observatory offers a model for responsible monitoring beyond Open Science itself. In doing so, it strengthens the evidence base for a more open, inclusive, and sustainable European Research Area [2].

This project has received funding from the European Union’s Horizon Europe framework programme under grant agreement No. 101148217. 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 Research Executive Agency. Neither the European Union nor the European Research Executive Agency can be held responsible for them.

#### Links:

[L1] <https://open-science-monitoring.org/principles/>

[L2] <https://www.eoscobservatory.eu/>

[L3] <https://graph.openaire.eu/>

#### References:

- [1] G. O’Neill, “Monitoring Framework for National Contributions to EOSC and Open Science”, 2025. DOI: 10.5281/zenodo.14919248.
- [2] T. Szybisty, “Monitoring Open Science across ERA: Early Lessons from the EOSC Open Science Observatory”, 2025. DOI: 10.5281/zenodo.16991651.

#### Please contact:

Tereza Szybisty  
OpenAIRE Amke, Czech Republic  
[tereza.szybisty@openaire.eu](mailto:tereza.szybisty@openaire.eu)

## From Data to Knowledge: Open Pipelines for Collaborative Production Research

by Leon Gorissen, (RWTH Aachen University) and Philipp Brauner (RWTH Aachen University) Germany)

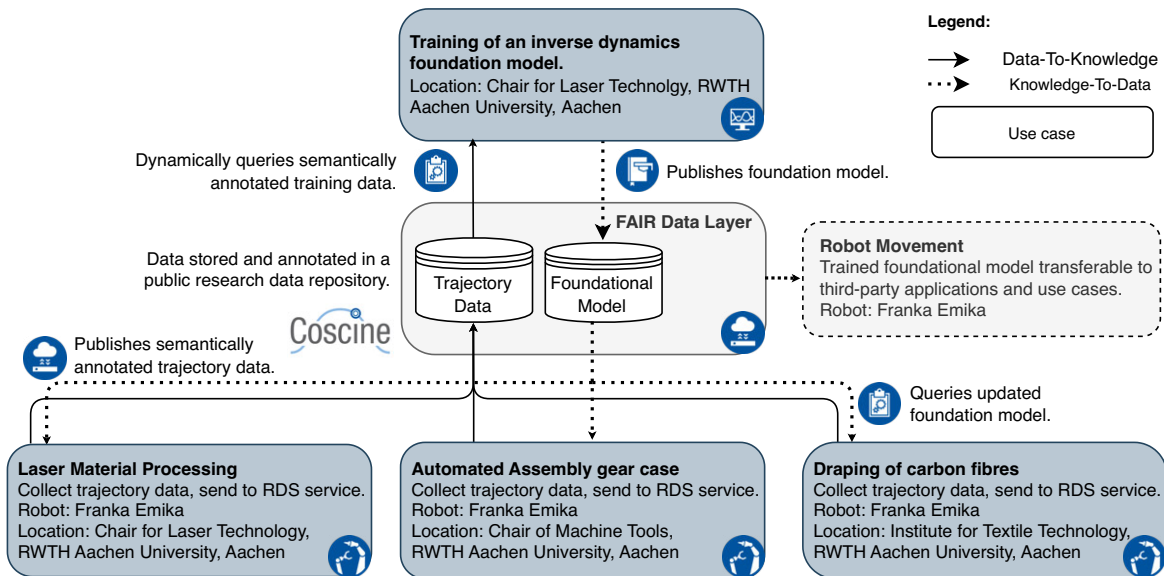
*Open Science promises better, faster, and more trustworthy research and benefits arise if data can be shared, aggregated, and reused. We present data-to-knowledge pipelines in an engineering context as a practical way to turn open research data into additional insights and tangible benefits across organizations. Drawing on shared data from industrial robots across organisations, the work illustrates how open infrastructures can enable collaboration at scale.*

Open Science has become a strategic priority for the European Commission and other funding agencies, with the aim of making research more transparent, reusable, and collaborative. A key element of this vision is the effective use of open data: not only by publishing datasets, but by enabling data to be combined, interpreted, and transformed into knowledge across disciplines and institutions. We present a framework that addresses this challenge through data-to-knowledge (D2K) pipelines (Figure 1), providing a practical approach to implementing Open Science in a complex, collaborative research cluster focused on the digital transformation of production [1].

The article shows how openly shared research data from different organisations can be systematically transformed into insights that benefit multiple stakeholders. Instead of treating data publication as an end point, the proposed pipelines emphasize continuous processes in which data is produced, shared, enriched with meaning, analysed, and then fed back to inform new experiments, models, or decisions. This approach aligns with the principles of Open Science, particularly reuse, transparency, and reproducibility.

A central concept in the work is the use of Digital Shadows [1]. These extend the idea of Digital Twins by providing context- and task-dependent, purpose-driven representations of real-world entities such as machines, processes, or experiments. Digital Shadows capture both data and its contextual meaning. By standardizing how data is described and linked to real-world objects, they make shared data easier to understand and reuse, even beyond its original application. This is particularly important for interdisciplinary research, where data often needs to move across domain boundaries.

We demonstrate the data-to-knowledge pipelines through an industrial case study involving robotic production systems operated at different sites and by different organisations. Trajectory data from identical robot models, used independently by each organisation, is shared through a common open research data infrastructure (the open source framework Coscine [L1]). Because the data is described in a consistent, reusable way and follows the FAIR principles, it can be easily



**Figure 1: Data-to-knowledge pipeline illustrating how shared production data are transformed into reusable models and insights across organisations.**

identified and queried by other processes. This enables an independent second process to query the shared data and train a more robust base model for robot control, benefiting from increasingly diverse trajectory data aggregated across sites without additional effort. The resulting model is equally openly available and improves robotic operation across sites, illustrating how open data can create collective benefits without requiring full centralisation of resources or expertise.

Importantly, the framework respects organizational autonomy while enabling collaboration. Each participant retains control over its own data and systems, while contributing selected data to a shared environment under mutually agreed conditions. This federated approach represents a highly relevant model for Open Science infrastructures, where legal, ethical, or commercial constraints often preclude full data openness. In this setting, data sharing yields reciprocal benefits: contributors and the broader community gain access to a more robust shared base model for robot control, which can in turn enhance local operations while supporting collective advancement. More broadly, the framework enables a “world-wide lab” (WWL) paradigm in which production environments function as distributed experiments, contributing empirical evidence toward a deeper understanding of factors influencing performance, quality, and safety [1,2]. Crucially, participation requires only limited additional effort while offering tangible gains through collaboration.

From an Open Science perspective, the proposed data-to-knowledge pipelines support several key goals. They improve reproducibility by making data findable, accessible, and interoperable, and reusable as data is publicly available and transformations explicit and traceable. They enhance reuse by providing semantic descriptions that allow data to be understood beyond its original project. They also encourage collaboration, as shared pipelines and models create incentives for participants to contribute data and expertise.

The article also looks ahead to future research infrastructures. As scientific data volumes continue to grow, simply storing

open datasets will not be enough. Especially in the domain of resource intensive training of machine learning models, researchers will need shared tools, pipelines, and services that help turn data into knowledge efficiently and responsibly. The presented framework offers a blueprint for such infrastructures, building on open standards and tools, using modular components, and demonstrating real-world applicability.

In summary, the work shows how Open Science can move beyond ideals toward implementation and applicability. By focusing on data-to-knowledge pipelines, it demonstrates how open data can become a living resource, continuously reused, enriched, and transformed, supporting transparent and collaborative research across disciplines and sectors.

**Acknowledgements:** Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy — EXC-2023 Internet of Production — 390621612.

**Link:**

[L1] <https://coscine.rwth-aachen.de/>

**References:**

- [1] P. Brauner et al., “A Computer Science Perspective on Digital Transformation in Production,” *ACM Trans. Internet Things*, vol. 3, no. 2, 2022, doi: 10.1145/3502265.
- [2] M. Behery et al., “Vision Paper: Leveraging Industrial Big Data – Past, Present, and Future of the World Wide Lab,” in *Proc. IEEE Int. Conf. on Big Data*, 2023, doi: 10.1109/BigData59044.2023.10386257.
- [3] L. Gorißen et al., “Demonstrating Data-to-Knowledge Pipelines for Connecting Production Sites in the World Wide Lab,” , *arXiv preprint arXiv:2412.12231*, 2024. [Online]. Available: <https://arxiv.org/pdf/2412.12231>.

**Please contact:**

Leon Gorißen, Chair for Laser Technology  
RWTH Aachen University, Germany  
[leon.gorissen@ilt.rwth-aachen.de](mailto:leon.gorissen@ilt.rwth-aachen.de)

# Scalable Platforms for Reusing Research Infrastructure Data in Open Science

by Magdalena Brus (EGI Foundation), Ville Tenhunen (EGI Foundation), Gergely Sipos (EGI Foundation), on behalf of the RI-SCALE consortium

**European Research Infrastructures generate vast amounts of valuable scientific data, yet reusing these data remains difficult in practice due to the growing gap between data availability and data usability. As datasets reach terabyte and petabyte scale, downloading, storing, and analysing them locally becomes impractical, while configuring suitable computational and AI environments presents significant barriers. This article presents a new platform-based approach that brings data, computation, and artificial intelligence tools and applications together, enabling more effective, transparent, and scalable Open Science.**

Research Infrastructures (RIs) play a central role in the European research ecosystem. They produce high-quality, curated datasets that underpin scientific progress in domains such as climate science, atmospheric physics, biomedicine, and imaging. Advances in sensors, cameras, and digital instrumentation have dramatically increased both the volume and complexity of this data, turning many RIs into large-scale data producers.

In parallel, Open Science has become a strategic priority in Europe. Policies promoting open access, FAIR data principles, and reproducible research have significantly improved the availability of research data. However, open availability alone does not guarantee effective reuse. As datasets grow to terabyte and petabyte scale, researchers increasingly face practi-

cally downloading large datasets to local machines is often infeasible, while configuring suitable analysis or AI environments remains a major obstacle for many users.

This growing gap between open data and usable data limits the societal and scientific value of RI data holdings. While some RIs offer tailored analysis services or cloud-based environments, these solutions are often domain-specific, difficult to scale, or costly to maintain. A more general, interoperable approach is needed to support Open Science across disciplines and infrastructures.

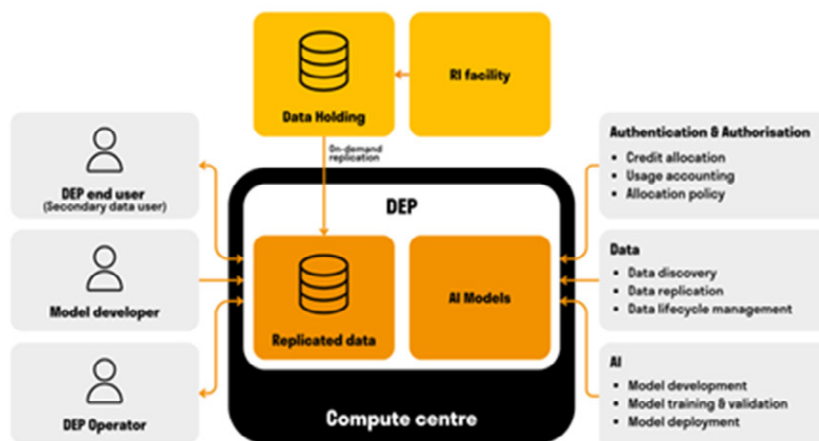
Recent technological developments provide new opportunities to address this challenge. Cloud and high-performance computing infrastructures, containerisation technologies, interactive notebooks, and federated identity management now make it possible to analyse data where they are hosted rather than moving data to users. At the same time, AI has become a key driver of scientific discovery, enabling new forms of data analysis, pattern detection, and automation across many research domains.

Against this backdrop, the EU-funded RI-SCALE project introduces the concept of Data Exploitation Platforms as a new way to enable scalable reuse of RI data in Open Science. A Data Exploitation Platform (Figure 1) extends an RI's data holdings with co-provisioned computational services, allowing users to analyse data and develop AI applications directly on scalable computing infrastructures. Rather than replacing existing repositories, these platforms complement them by transforming static data collections into active environments for exploration and reuse.

Data Exploitation Platforms are designed as open, modular systems built on open-source software and open standards. They integrate three main functional elements. First, data lifecycle management services enable trusted replication, caching, and orchestration of datasets from RI repositories onto suitable compute resources. This reduces data movement, improves performance, and supports transparent provenance tracking. Second, computational environments provide ready-

## RI SCALE Solution: Data Exploitation Platform

A Scalable, Open Platform for Scientific Data Valorisation



cal barriers related to computing capacity, software environments, and data handling expertise. Downloading large

to-use tools such as interactive notebooks, data analytics frameworks, and AI toolkits, lowering technical barriers for users. Third, trust and access management mechanisms support federated authentication and authorisation, enabling both open and controlled access to data and computing resources in line with legal, ethical, and policy requirements.

By embedding AI frameworks and reusable models directly within the data environment, these platforms support AI-driven Open Science. Researchers can reuse existing models, adapt community-developed solutions, or train new models on large RI datasets without needing to manage complex infrastructure. At the same time, RI operators can apply AI techniques to improve data quality, enhance metadata, and support FAIRification processes. Importantly, provenance information for both data and models can be preserved, supporting transparency and reproducibility.

The Data Exploitation Platform approach is being validated across several European Research Infrastructures in environmental and health sciences. These include domains such as climate modelling, atmospheric research, biobanking, and bioimaging, where data volumes and reuse potential are particularly high. The platforms are also designed to interoperate with emerging European Data Spaces, enabling enrichment of RI data with external sources and supporting cross-domain research while respecting data protection and security constraints.

Beyond technical capabilities, sustainable Open Science depends on skills, governance, and long-term operational models. For this reason, the platform concept is complemented by competence centres and training activities that support researchers, RI staff, and external stakeholders in adopting scalable data and AI solutions. These activities promote knowledge sharing, responsible use of computational resources, and awareness of energy efficiency and environmental impact.

Looking ahead, Data Exploitation Platforms offer a concrete pathway toward more effective Open Science infrastructures in Europe. By bringing computation to data and integrating AI capabilities into trusted, interoperable environments, they help unlock the full value of publicly funded research data. In doing so, they support a transition from data availability to data usability, accelerating discovery and innovation for the benefit of science and society.

#### Links:

[L1] RI-SCALE project website: <https://www.riscale.eu/>

[L2] RI-SCALE use cases: <https://www.riscale.eu/use-cases>

#### Please contact:

Magdalena Brus

EGI Foundation

[magdalena.brus@egi.eu](mailto:magdalena.brus@egi.eu)

## MOEBA-BIO: An Open and Extensible Framework for Evolutionary Biclustering in Biomedicine

by Adrián Segura Ortiz, José García-Nieto (ITIS Software, University of Málaga) and Laetitia Jourdan (CRISAL Laboratory, University of Lille)

*MOEBA-BIO is an open and extensible evolutionary biclustering framework that empowers the biomedical research community to specialize in advanced optimisation strategies. Through its modular design, holistic representation of solutions, adaptive learning of the number of biclusters, and integrated self-configuration mechanisms, the framework enables expert-driven domain refinement while promoting sustainable open research software.*

#### From Algorithm to Open Research Software

Open Science increasingly requires that computational methods are not only published, but shared as reusable, extensible and sustainable research artefacts. In computational biomedicine, however, advanced optimisation techniques are often released as standalone algorithms tightly coupled to specific studies. While scientifically valuable, such implementations can be difficult to extend, adapt to new domains, or integrate into broader research infrastructures.

Transforming algorithmic contributions into open research software requires architectural decisions that prioritise modularity, interoperability and long-term maintainability from the outset. In this context, researchers at the University of Málaga (Spain), within the Khaos Research Group [L1] and the ITIS Software Institute [L2], developed MOEBA-BIO [1] in close collaboration with the ORKAD research team [L3] of the CRISAL laboratory [L4], University of Lille (France), as an open framework for evolutionary biclustering in biomedical data analysis. Rather than delivering a single method, it provides a configurable infrastructure in which representations, objective functions, evolutionary operators and optimisation strategies are separated into independent components. This modular architecture allows researchers to assemble, extend and refine evolutionary biclustering strategies without re-designing the entire system. The overall architecture of the framework and its modular objective space are illustrated in Figure 1.

#### A Holistic Representation for Structural Learning

A central design decision of the framework is its holistic representation of solutions. Instead of encoding individual biclusters independently, each candidate solution models a complete structure. This shift enables evaluation strategies that consider not only the internal quality of each bicluster, but also the global organisation and coherence of the solution as a whole.

This global perspective makes it possible to design objective functions that evaluate solution-wide properties, such as differentiation between groups, overall coverage or coherence

with biological constraints. Rather than optimising isolated substructures, the framework supports modelling decisions that reflect the systemic nature of biomedical data. Evolutionary search thus becomes structurally aware instead of purely local.

### Adaptive Learning of the Number of Biclusters

One direct consequence of this holistic representation is the adaptive determination of the number of biclusters. Traditional biclustering workflows often require this parameter to be fixed in advance or selected through external heuristics [2]. In MOEBA-BIO, however, the quantity of biclusters evolves dynamically as part of the optimisation process.

By embedding solution size into the evolutionary encoding, model complexity becomes an optimisable dimension. The algorithm can determine how many coherent groups best describe the data under the selected objectives, without imposing artificial constraints. This is particularly relevant in biomedical applications, where predefined structural assumptions may bias interpretation or overlook latent biological organisation.

This capability reduces the need for post-processing or manual adjustment and promotes methodological transparency. Because the number of groups is learned under the same optimisation criteria that guide structural coherence and domain-specific objectives, the resulting models remain internally consistent.

### Supervised and Unsupervised Self-Configuration

Beyond structural learning, MOEBA-BIO incorporates a dedicated self-configuration mechanism designed to facilitate domain-driven specialisation. This component operates in two complementary phases.

In a supervised configuration phase, the framework evaluates alternative combinations of objectives and algorithmic components against representative benchmark datasets for which reference structures are available. Using wrapper-based optimisation, it identifies configurations that best align optimisation behaviour with domain expectations. This allows the framework to automatically select and weight objective functions in a way that reflects the characteristics of a specific biomedical context.

Once this higher-level configuration is established, an unsupervised refinement phase further tunes optimisation parameters such as evolutionary operators and algorithm settings. This second stage focuses on improving convergence and solution quality without relying on external labels, ensuring adaptability to new datasets while preserving generality.

Together, these supervised and unsupervised layers transform the framework into an adaptive analytical environment. Domain experts can guide specialisation through representative data, while the internal optimisation process refines persistent

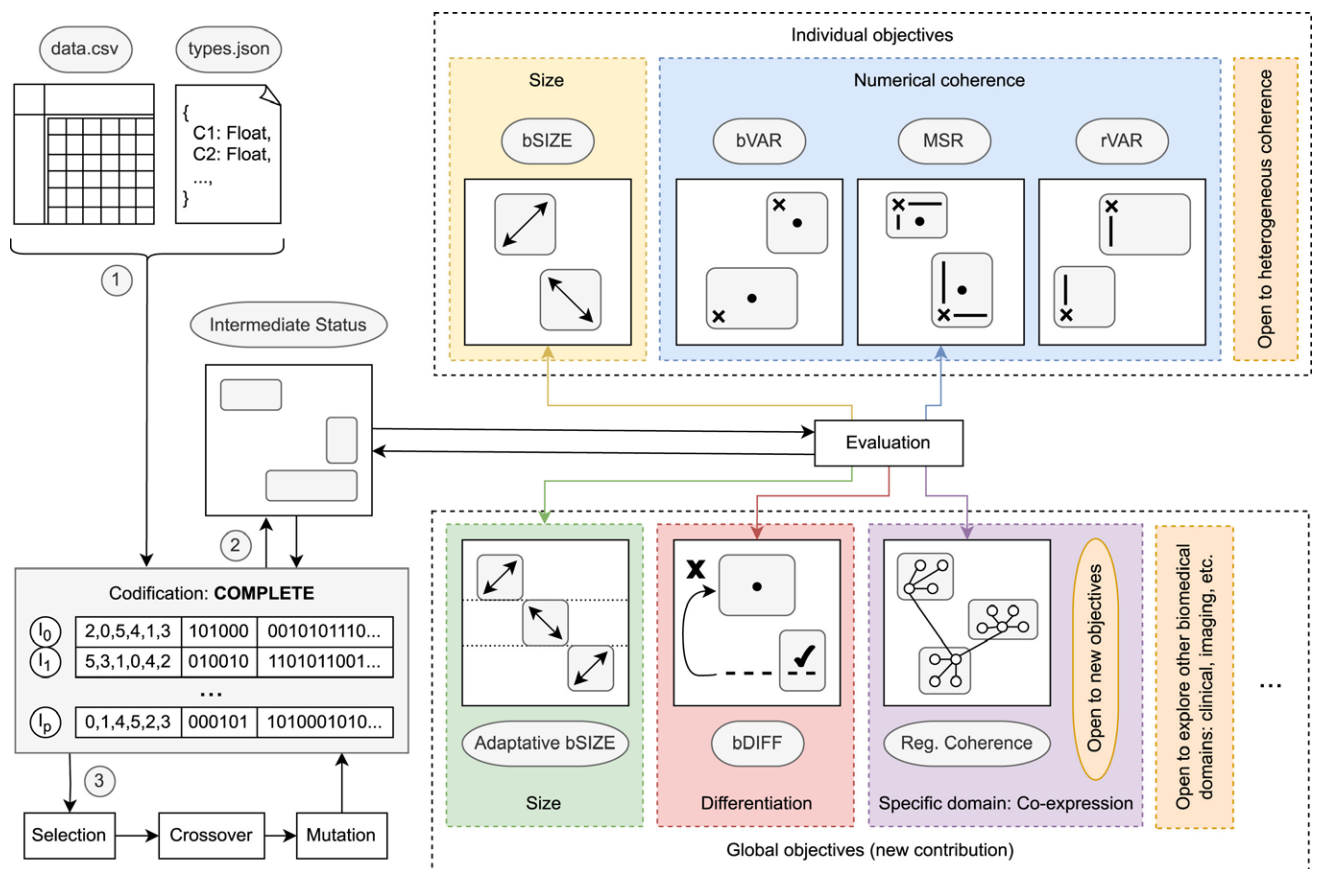


Figure 1: Overview of the MOEBA-BIO framework architecture. The workflow begins with structured biomedical input data and a column-type specification, which are transformed into a holistic solution representation (“COMPLETE” encoding). Each individual models a full biclustering structure and evolves through selection, crossover and mutation. The evaluation stage integrates traditional individual objectives (e.g., size and numerical coherence) together with newly introduced global objectives such as adaptive bicluster size, structural differentiation and domain-specific regulatory coherence. The modular design allows the integration of additional objectives and the exploration of heterogeneous biomedical domains, supporting extensibility and expert-driven specialisation within an open research software ecosystem.

formance autonomously. This combination reduces technical barriers and supports expert-driven domain refinement without sacrificing reproducibility.

#### Open, Interoperable and Extensible by Design

MOEBA-BIO was developed following open research software principles. The framework is fully open source, documented and publicly available through GitHub [L5]. Its architecture encourages extension: new objective functions, representations or evolutionary operators can be integrated without modifying the core system.

To promote interoperability, the framework is distributed through the Maven Central Repository [L6], enabling seamless integration into Java-based research pipelines and larger analytical infrastructures. This design choice supports sustainable reuse, reproducible workflows and incorporation into educational environments.

By clearly separating representation, evaluation and optimisation layers, the framework enhances transparency and traceability. Researchers can inspect, modify and extend each component independently, strengthening methodological openness.

#### Demonstrated Specialisation in Gene Co-Expression

The extensibility of the framework has already been demonstrated in gene co-expression analysis. By incorporating a regulatory coherence objective derived from inferred gene regulatory networks [3], MOEBA-BIO was specialised to align biclustering results with biologically meaningful modular structures.

Importantly, this domain adaptation did not require redesigning the system. Instead, new objectives were integrated into the existing architecture, illustrating how the framework enables community-driven extension rather than isolated reimplementation.

#### Towards Sustainable Evolutionary Research Infrastructures

As biomedical datasets grow in complexity and heterogeneity, analytical tools must evolve beyond fixed algorithms. Open frameworks that enable structural learning, adaptive configuration and domain-driven specialisation are essential for sustainable scientific progress.

MOEBA-BIO illustrates how evolutionary optimisation can transition from isolated methodological contributions to reusable, extensible research infrastructure. By embedding modularity, interoperability and expert-guided adaptability into its core design, the framework aligns with broader Open Science priorities centred on transparency, sustainability and collaborative development.

#### Links:

[L1] <https://khaos.uma.es/>

[L2] <https://itis.uma.es/>

[L3] <https://orkad.univ-lille.fr/>

[L4] <https://www.cristal.univ-lille.fr/>

[L5] <https://github.com/AdrianSeguraOrtiz/MOEBA-BIO>

[L6] <https://kwz.me/hIP>

#### References:

- [1] A. Segura-Ortiz, et al., “Exhaustive biclustering driven by self-learning evolutionary approach for biomedical data,” *Computer Methods and Programs in Biomedicine*, p. 108846, 2025.
- [2] M. Vandromme, et al., “A biclustering method for heterogeneous and temporal medical data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 2, pp. 506–518, 2020.
- [3] A. Segura-Ortiz, et al., “Multifaceted evolution focused on maximal exploitation of domain knowledge for the consensus inference of Gene Regulatory Networks,” *Computers in Biology and Medicine*, vol. 196, p. 110632, 2025.

#### Please contact:

Adrián Segura Ortiz

ITIS Software, University of Málaga, Spain

[adrianseor.99@uma.es](mailto:adrianseor.99@uma.es)

José Francisco Aldana Montes

ITIS Software, University of Málaga, Spain

[jfaldana@uma.es](mailto:jfaldana@uma.es)

## Reinforcing Open Science in Biodiversity through Semantic Knowledge Graphs

by Yannis Marketakis, Eleni Tsouloucha, Athina Kritsotaki, and Yannis Tzitzikas (FORTH-ICS)

***Transforming FishBase into a semantic knowledge graph shows how legacy biodiversity databases can become FAIR, interoperable infrastructures that enable reuse and integration in Open Science.***

Open Science is one of the cornerstones of European research policy, promoting transparency, accessibility, reuse, and collaboration across scientific domains. While many scientific communities have made progress in opening data, experience shows that availability alone does not guarantee meaningful integration, reuse, or long-term sustainability. Addressing this challenge requires data infrastructures that go beyond simple access, supporting semantic integration, evolution, and machine-readable access, enabling data to be combined, queried, and reused across domains.

Biodiversity research provides a compelling paradigm for these challenges. FishBase, one of the most widely used global databases on fish species, has long played a crucial role in supporting research, education, and policy. Its contents are openly accessible and continuously curated, making it an important resource for the biodiversity community. Despite its importance, like many other legacy databases, FishBase was originally designed as a standalone information system. As a result, integrating its data with other research resources, performing complex cross-domain queries, or reusing its content in new contexts often requires manual effort.

This experience highlights a broader Open Science challenge: how to support the evolution of well-adopted, widely used open databases into FAIR-by-design data infrastructures. In what follows, FishBase is used as a concrete example to illustrate how semantic knowledge graphs, open APIs, and evolution workflows can reinforce Open Science by enabling interoperability, reuse, and long-term sustainability of biodiversity data, as demonstrated in recent Open Science efforts around biodiversity knowledge graphs [1]. Figure 1 provides an overview of this transformation and its role in enabling integration with external Open Science infrastructures.

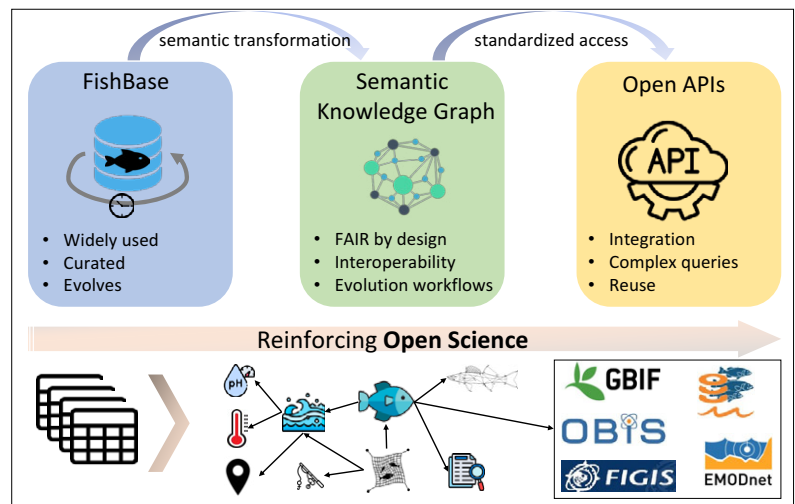
Applying FAIR data principles in practice requires more than just making data available online. It requires data to be described in a structured and semantically explicit manner, accessed through standardized interfaces, and supported by mechanisms that ensure consistency and sustainability as data evolve over time. This also directly supports reproducibility and transparency, since datasets, structures, and relationships can be inspected, queried and recombined in a verifiable and machine-interpretable manner. Thus, FAIR data management is not an afterthought, but a design choice that must be embedded into the data infrastructure and its governance policy.

The experience with FishBase clearly illustrates these requirements. Although FishBase has long been an open and continuously curated resource, its original design as a conventional database limits cross-domain integration and large-scale reuse. Integrating its data with external biodiversity resources, answering complex queries that require combining different pieces of information, or supporting data reuse typically requires manual effort.

The transformation of FishBase into a semantic knowledge graph, carried out within the SemantyFish initiative [1] [L1], addresses these limitations by providing a FAIR-by-design representation of biodiversity data. Species, biological characteristics, ecology, population dynamics, life cycle and history, distribution, and other resources are modelled as semantically defined entities connected through explicit relationships. This makes data machine-interpretable and easier to integrate across domains. To this end, ontologies play a central role, offering shared conceptual models that ensure semantic consistency and interoperability with other biodiversity infrastructures.

Furthermore, access and reuse are further enhanced through open APIs, which expose the semantic knowledge graph via standardized, programmatic interfaces. This layer enables both human-oriented applications and machine-operated services and workflows to access up-to-date data in a structured, standardized, and reusable form, without relying on static exports or other ad hoc solutions.

Finally, evolution workflows ensure that the semantic representation remains aligned with updates to the original FishBase database, without requiring changes to FishBase's existing data management or update policies. As new data are added or existing data are revised, changes are systematically propagated to the knowledge graph, preserving accuracy and



**Figure 1:** The transformation workflow of FishBase from a relational database into a FAIR-by-design semantic knowledge graph, supported by evolution workflows and open APIs enabling interoperable access and integration with external Open Science infrastructures.

trustworthiness over time. Overall, semantic modeling, knowledge graph construction, open APIs, and evolution workflows demonstrate how a widely used biodiversity database can be transformed into a FAIR, sustainable infrastructure that supports Open Science.

This transformation enables forms of sharing and reuse that were previously difficult to achieve. The explicit representation of entities and relationships allows researchers to formulate complex queries that traverse species, ecological traits, habitats, and references, supporting biodiversity research and analyses that go beyond predefined database views. At the same time, machine-interpretable semantic resources make it possible to programmatically combine FishBase data with external datasets, facilitating automated workflows and interoperability, as demonstrated in cross-domain knowledge graph applications for sustainable aquafood communication [2].

The experience with FishBase illustrates how rethinking data infrastructures can strengthen Open Science practices. Designing for interoperability, reuse, and evolution of existing scientific data resources proved essential for enabling meaningful data sharing without disrupting established practices. Looking ahead, such approaches can support broader perspectives, including cross-domain research by providing sustainable foundations for Open Science.

**Link:** [L1] <https://semantyfish.github.io/>

#### References:

- [1] Connecting fish data for open and sustainable science. EuroFish Magazine, Issue 6, 2025. <https://eurofish.dk/connecting-fish-data-for-open-and-sustainable-science/>
- [2] Y. Marketakis, et al “Construction, Access and Application of a Knowledge Base for Sustainable Aquafood Communication”, 14th International Joint Conference on Knowledge Graphs. Heraklion, Greece, 2025.

#### Please contact:

Yannis Marketakis, FORTH-ICS, Greece  
marketak@ics.forth.gr

# Unlocking Machine-Composability with the EOSC Interoperability Framework Registry

by Alessia Bardi (CNR-ISTI, OpenAIRE AMKE), Konstantina Galouni (ARC) and Paolo Manghi (CNR-ISTI, OpenAIRE AMKE)

*The EOSC Beyond project is transforming the European Open Science Cloud into a dynamic, programmable environment by evolving the Interoperability Framework Registry from human-readable documents into machine-actionable templates. This shift empowers the scientific community to programmatically compose resources into automated scientific workflows.*

The European Open Science Cloud (EOSC) is a federation of thematic and national research infrastructures with the ambition to develop a web of FAIR data and services for data-driven, Open Science. Despite the availability of a large volume of research resources in the federation, the composition of these diverse resources into integrated scientific workflows has historically remained a human-resource intensive process. To address this, the EOSC Beyond project [L1] is advancing the EOSC to support machine-composability, allowing tools and services to combine resources programmatically [1]. Central to this shift is the evolution of the EOSC Interoperability Guidelines (IF) Registry. An Interoperability Guideline (IG) is defined as a set of instructions that EOSC Providers must implement to enable specific functionalities and ensure their services or research products can work to-

gether seamlessly within the EOSC ecosystem. For example, there are interoperability guidelines for onboarding research products (metadata exchange) and for compliance with the EOSC Monitoring Service.

## From Human-Readable Guidelines to Machine-Actionable Templates

In the current state of the art, the EOSC Interoperability Framework (IF) Registry primarily supports the registration of human-readable guidelines. While these were essential for human developers, they lacked the structure required for automated deployment or systemic orchestration.

The next generation of the IF Registry evolves into a collection of guidelines that support the programmatic composition of resources. This is achieved through two new technical pillars:

- **Configuration Templates:** These are structured metadata profiles that allow providers of guidelines to detail which information must be provided by a resource (e.g. a service) that is compliant with the guideline. For example, the IG for onboarding research products require the service to specify the base URL of its OAI-PMH endpoint and other details required for the harvesting.
- **Configuration Instances:** These are the specific, populated parameters provided by a service that indicate how it complies with a given template. For example, if the provider registers a service and declares it is compliant with the IG for onboarding research products, then it must provide the actual OAI-PMH endpoint, and the other required details, such as the metadata format it supports.

By recording these machine-actionable descriptions in the EOSC Service Registry, the framework enables core components and third-party applications to interact with services without manual intervention.

**EOSC Metadata Harvester** | **eosc** | **BEYOND**

Filters

Compatibility (6)

- openaire2.0\_data
- openaire4.0
- driver
- a
- test
- UNKNOWN

14 Services, Page 1 of 2

**B2FIND**

Web page: <https://www.eudat.eu/services/b2find><sup>12</sup>  
Base URL: <http://b2find.eudat.eu/oai><sup>12</sup>  
Compatibilities: openaire2.0\_data, openaire4.0

B2FIND is the EUDAT metadata indexing service and provides a discovery portal which allows users to find data collections within an international and inter-disciplinary scope. It is based on a comprehensive metadata catalogue of research data collections stored in EUDAT data centres and community repositories. Harmonization of the metadata descriptions collected from heterogeneous sources enables not only the presentation in a consistent form but as well the faceted search across scientific domain boundaries. For Communities and other providers of research data who need to publish and give visibility to their metadata and individual researchers who need to search data from everywhere, and see data in the context with an across community approach. ••Features•• - Harmonization of the metadata descriptions via the EUDAT Core metadata schema - Harvesting of repositories via different protocols (e.g. OAI-PMH, CSW, Rest-APIs) - Faceted search via 12 facets (including geospatial and temporal search options), additional free text search - Metadata aggregation from community repositories, multiple metadata standards are supported - Harvested by OpenAIRE explorer

HARVEST →

**Zenodo**

Web page: <https://www.zenodo.org><sup>12</sup>  
Base URL: <https://zenodo.org/oai2d><sup>12</sup>  
Compatibilities: openaire2.0\_data

Zenodo is a general purpose repository that enables researchers, scientists, projects and institutions to share, preserve and showcase multidisciplinary research results (data, software and publications) that are not part of the existing institutional or subject-based repositories of the research communities. It is founded in the trustworthy CERN data centre. Enables everyone to participate in Open Science. Used by more than 50K researchers and 3K communities all over the world.

HARVEST →

**ROHub**

Web page: <https://elliance.rohub.org/><sup>12</sup>  
Base URL: <https://api.rohub.org/api/oai2d/><sup>12</sup>  
Compatibilities: openaire4.0

ROHub enables the management, sharing and preservation of research artefacts related to a particular investigation/research challenge as a single information unit through the

Figure 1: The Scholarly Communication Metadata Aggregator demo application.

### Case Study: The Scholarly Communication Metadata Aggregator

The power of this machine-actionable approach is best demonstrated by the Scholarly Communication Metadata Aggregator for EOSC Nodes [L2]. This application highlights how machine-readable guidelines can automate scholarly communication workflows that previously required human intervention.

In a traditional setup, a user would need to manually find a repository, identify its OAI-PMH endpoint, and configure a harvesting tool. Under the new framework, the process is fully automated:

1. **Discovery:** The application queries the Service Registry to identify data sources that are compliant with the Interoperability Guidelines for onboarding research products (see Figure 1).
2. **Automated Configuration:** Instead of manual setup, the application reads the Configuration Instance of the selected service, which contains the exact parameters needed to harvest its metadata.
3. **Execution:** The application launches the harvesting process automatically based on these machine-read parameters.
4. **Delivery:** Once the process is complete, the user is automatically notified via email with a link to download the aggregated metadata records.

### Conclusion and Future Impact

EOSC Beyond is creating a "Web of FAIR Data and Services" that is composable in automated workflows. The framework reduces the technical burden on researchers and service providers alike, allowing them to focus on research rather than integration code. The Scholarly Communication Metadata Aggregator demonstrates the potential of machine-actionable interoperability guidelines for the realisation of automated or semi-automated workflows. The EOSC Beyond technical roadmap includes further enhancements both to the framework and the use case application to improve the user experience and integrate other services of the EOSC ecosystem, such as the EGI Data Transfer for safe transfer of harvested records in a remote location selected by the user.

**Acknowledgment:** This work is partly founded by the European Union EOSC Beyond project grant agreement 101131875. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

### Links:

[L1] <https://www.eosc-beyond.eu/>

[L2] <https://kwz.me/hIg>

### Reference:

- [1] Bardi, A., Brandt, C., Caballer, M., Moltó, G., & Mantes, T. (2025). EOSC Beyond D13.2 First report on EOSC Execution Framework Architecture and capabilities. Zenodo. <https://doi.org/10.5281/zenodo.17234674>

### Please contact:

Alessia Bardi, CNR-ISTI, OpenAIRE, Italy  
[alessia.bardi@isti.cnr.it](mailto:alessia.bardi@isti.cnr.it)

## Human-Centered Threat Modeling in Web Standardization at W3C

by Simone Onofri (W3C) and Giovanni Corti (FBK)

*W3C has published a Threat Modeling Guide that expands security analysis to include societal impacts and promotes the publication of reusable, public threat models, aligning threat modeling with Open Science principles of transparency, reuse, and accountability.*

Threat modeling is commonly described as a structured practice for defining a system, identifying threats and potential attacks, selecting responses to address them, and iterating as the system evolves [1]. In standardization, the system is often the specification, and how it is implemented and deployed in the environment. When a specification forms part of the foundation of a digital infrastructure, the threat model should not only capture threats to the system but also threats generated by the system for people, communities, and the environment. This perspective aligns with Open Science by emphasising transparency in design decisions and making the societal impacts of digital infrastructure visible and accountable.

### From a system-centric to a human-centric threat modeling

The Threat Modeling Guide integrates the stakeholder perspective. After the classic first step, which answers the question “what are we working on” [2], and then after modeling the system, it also asks “who is impacted” by the system, conducting a stakeholder impact analysis, where we consider not only the users of the system, but also those who are not users, for example because they may have been excluded by the design of the system itself. The goal of this step is to identify these stakeholders to discuss design trade-offs as early as possible and provide mitigations at the standard level. By making stakeholder impacts explicit, this approach supports more transparent and inclusive design processes, which are central to Open Science practices.

### From threats to the system to threats to humans

After identifying the system and its stakeholders, the next step is still “what can go wrong,” but with a broader scope: to understand both threats to the system and threats generated by the system for its stakeholders. This is a crucial phase and brings a well-known challenge in security threat modeling: developers with expertise in the system may lack security expertise, while security experts knowledgeable about threats and impacts may lack expertise in the specific system.

This challenge is even more pronounced when the terminology and the vocabulary used to identify societal impacts, which are expressed using aspirational or symbolic language (e.g., fairness, dignity, non-discrimination), must be translated into concrete technical and governance levers.

To overcome this challenge, threat modeling practices are complemented by facilitation techniques, such as serious games, including card games, to assist in threat elicitation. For example, in addition to using documentation such as the “Self-

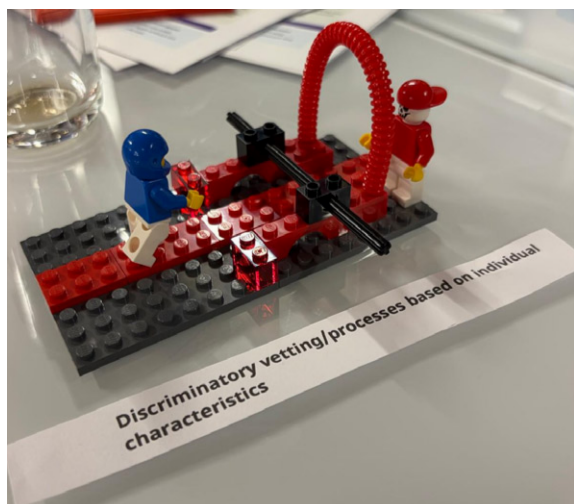


Figure 1: A model to illustrate discriminatory vetting/processes based on individual characteristics.

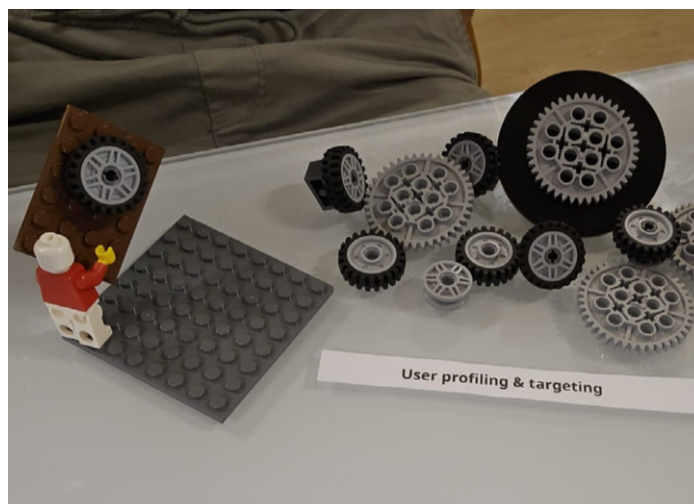


Figure 2: A model to illustrate user profiling and targeting.

Review Questionnaire: Security and Privacy” to identify issues, the W3C community also uses STRIDE cards for security, and LINDDUN cards for privacy, which are promising tools for helping even those unfamiliar with specific threats identify them. Addressing this gap is essential not only for effective security design but also for ensuring that knowledge about risks and impacts can be shared and understood across communities.

#### Bridging vocabularies with LEGO® SERIOUS PLAY® workshop

Digital identities, particularly those issued by governments, represent a high-impact societal use case, as they can serve as a technology of inclusion or exclusion or promote surveillance.

In October 2025, W3C experimented an approach that was deliberately not limited to textual discussions: a facilitated workshop with LEGO® SERIOUS PLAY® (LSP) [L1]. The objective was to help participants make abstract and societal impacts more concrete.

During the workshop, LSP was used as a facilitation tool for the Threat Modeling session. Starting with the harm taxonomy proposed in “Enhancing National Digital Identity Systems” [3], participants were asked to build a physical, metaphorical representation of the derived threats using LEGO bricks, thereby creating a threat landscape and connecting the various threats. The three-dimensional models served as boundary objects: they enabled participants with different professional and cultural vocabularies to converge on a common understanding of what an abstract harm is, how it manifests, and which technical or governance assumptions enable it.

#### Why public threat models matter

A final element emphasised by the guide is the publication of the threat model, a practice closely aligned with Open Science. Non-public threat models create knowledge asymmetries and limit reuse; public threat models make assumptions visible, clarify trade-offs, and expose the boundaries of responsibility among specifications, implementations, deployments, and governance, providing an element of accountability. In this way, public threat models function as open research artefacts

that can be inspected, reused, and built upon by a wider community.

From a purely research perspective, having public, open threat models make them reusable knowledge artifacts. They can be compared with others, extended by other standardization bodies or policymakers, and studied longitudinally as technologies evolve. In this sense, human-centered threat modeling becomes not only an engineering practice but also a means of analyzing the societal impact of technology. In this sense, human-centered threat modeling becomes not only an engineering practice but also a contribution to Open Science, enabling transparent, reusable, and comparable knowledge about the societal impact of technology.

#### Links:

- [L1] <https://kwz.me/hIu>
- [L2] <https://kwz.me/hIL>
- [L3] <https://digitallibrary.un.org/record/4031373>
- [L4] <https://kwz.me/hII>

#### References:

- [1] World Wide Web Consortium (2026) Threat modeling guide. Group Note Draft <http://w3.org/TR/threat-modeling-guide/>
- [2] Shostack, A. (2014). Threat modeling: Designing for security. John Wiley & Sons
- [3] G. Corti, G. Sasseti, A. Sharif, R. Carbone, and S. Ranise, “Enhancing National Digital Identity Systems: A Framework for Institutional and Technical Harm Prevention Inspired by Microsoft’s Harms Modeling,” in Proc. 22nd Int. Conf. on Security and Cryptography (SECRYPT 2025), 2025, pp. 723–728.

#### Please contact:

Simone Onofri, W3C  
simone@w3.org

# Co-Creation as Infrastructure for Open Science Data Spaces

by Rita Stampfl (University of Applied Sciences Burgenland), Silke Palkovits-Rauter (University of Applied Sciences Burgenland)

**Large-scale European research projects increasingly depend on interoperable data spaces and AI-driven services. Yet technical integration alone does not ensure sustainable and trustworthy ecosystems. A structured co-creation model, implemented since 2023 in a European energy data initiative, demonstrates how stakeholder collaboration, governance design and semantic alignment can be embedded directly into project architecture. By formalising participation across iterative development cycles and integrating ethical reflection into technological innovation, the model reframes co-creation as an infrastructural component of Open Science rather than a supplementary engagement activity.**

Across Europe, data spaces are being built to support digital transformation in domains such as energy, health and smart cities. Technical interoperability, AI-enabled analytics and governance frameworks are progressing rapidly. Yet many large-scale research consortia encounter a recurring challenge: how to align heterogeneous stakeholders beyond consultation to achieve shared ownership of outcomes? Since 2023, a European energy research initiative has addressed this issue by embedding a structured co-creation process directly into its project architecture. Instead of treating participation as an auxiliary activity, the consortium established a dedicated co-creation process connecting technology development, governance design and domain-specific use cases in the energy sector. The resulting model offers a transferable blueprint for Open Science infrastructures.

The generic co-creation process model (Figure 1) operates across three interdependent dimensions: technology, governance and energy as domain. Within iterative cycles, stakeholders jointly define challenges, analyse data landscapes, harmonise terminology, design services and refine prototypes through structured feedback loops.

A key insight from interdisciplinary research is that participation is often equated with user-centred design, while real decision-making power remains within technical teams [1]. Empirical findings show that users are frequently involved at the beginning (requirements gathering) or at the end (testing), but rarely throughout continuous co-decision processes. The presented model therefore formalises engagement levels ranging from information and consultation to collaboration and empowerment, explicitly clarifying roles and responsibilities.

Methodologically, the process combines structured and creative formats. LEGO Serious Play workshops surface tacit knowledge and build shared mental models. Business Model Canvas sessions clarify value creation and governance implications. World Cafés and Ishikawa diagrams support systemic exploration. Online co-creation is facilitated via shared Miro boards, common data overviews and collaboratively developed use-case visualisations. A jointly maintained terminology wiki ensures semantic interoperability – a foundational requirement for functioning data spaces.

Beyond methodological integration, the model embeds ethical reflection as a structural component. Recent analyses of co-creation in technological contexts emphasise the importance of addressing power imbalances and distinguishing between “co-creating with” and merely “testing on” communities [2]. In data-driven ecosystems, especially when working with sensitive infrastructures, these distinctions are crucial. Governance, regulatory and data sovereignty aspects are therefore discussed directly within the co-creation workshops rather than being addressed retrospectively.

Inclusive participation is a central design principle of the model. Interdisciplinary research consortia often involve insti-

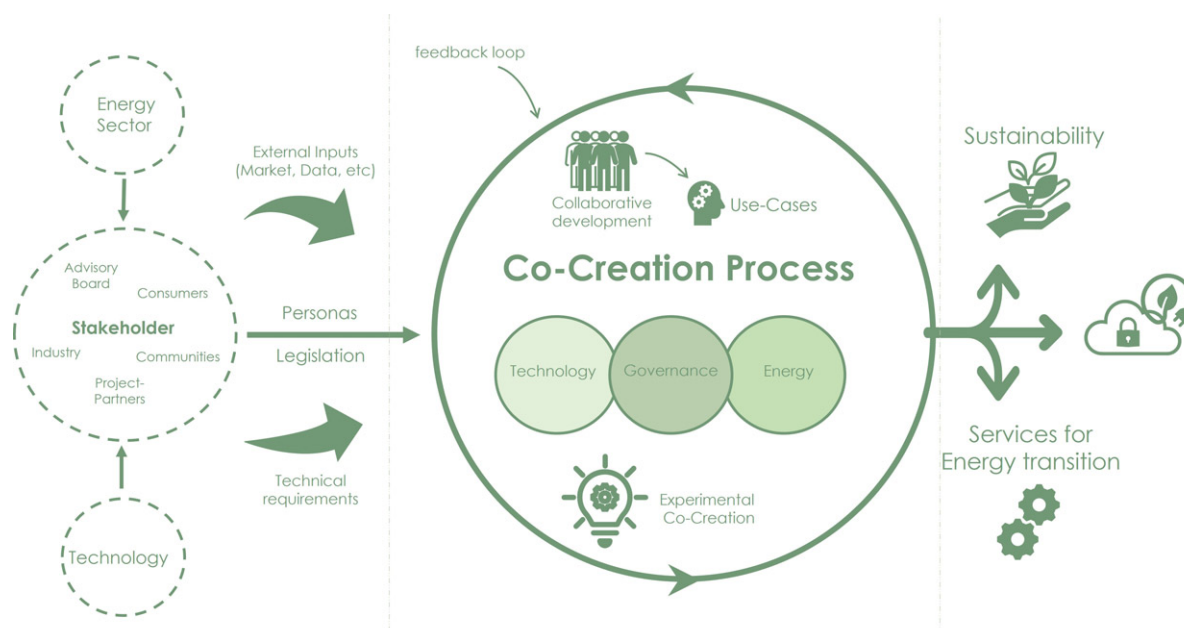


Figure 1: Generic Co-Creation Process Model connecting Technology, Governance and Domain through iterative stakeholder cycles.

tutional stakeholders but unintentionally exclude smaller actors, civil society organisations or less technically specialised participants. The structured co-creation workspace therefore deliberately lowers participation barriers by using visualisation techniques, shared vocabulary development and facilitated dialogue formats. Inclusion is understood not merely as representation but as the capacity to meaningfully influence decisions. By clarifying engagement levels and embedding empowerment mechanisms, the model addresses structural imbalances in knowledge, resources and digital literacy. In doing so, it aligns co-creation with broader European ambitions for equitable, citizen-centred Open Science ecosystems.

The approach has been implemented within the USEFLEDS initiative [L1], which develops an interoperable Energy Data Space to support sector-coupling flexibility in the energy transition. The project brings together Austrian research organisations, technology providers and industry stakeholders to design a data exchange model spanning the value chain of sustainable energy services.

Within this framework, cross-sector services such as flexible energy tariffs, energy flow optimisation and partial automation of ESG reporting have been co-designed. Importantly, these services emerged from structured stakeholder dialogue rather than purely technical specification. The co-creation process acted as a translational layer between AI-based analytics, governance requirements and operational energy practices.

This model contributes to Open Science in three ways. First, it enhances transparency by documenting shared artefacts and decision processes. Second, it supports FAIR-aligned data management principles by aligning terminology and usage conditions early in the development process [3]. Third, it strengthens trust by integrating ethical and legal considerations from the outset.

European initiatives such as GAIA-X [L2] highlight the importance of interoperability and digital sovereignty. However, technical standards alone do not guarantee sustainable ecosystems. The experience since 2023 suggests that co-creation itself must be considered part of the infrastructure: a socio-technical mechanism enabling shared understanding, reproducibility and long-term viability.

Future work focuses on formalising the model as a reusable framework for other domains and linking it with virtual research environments and Open Science infrastructures. Collaboration with additional European partners is planned to refine participatory standards and governance models for data-driven innovation. As Europe advances towards federated, AI-enabled and sovereign data spaces, structured co-creation may become a decisive factor for ensuring that technological interoperability is matched by institutional and societal alignment.

#### Links:

[L1] <https://usefleds.forschung-burgenland.at/645-2/>

[L2] <https://gaia-x.eu/>

[L3] <https://www.go-fair.org/fair-principles/>

#### References:

- [1] A Osterheider, et al., “Conceptualization of the Understanding of Participation and Co-Creation in Interdisciplinary Research Groups developing Digital Health Technology: An Exploratory Study: Conceptualization of the Understanding of Participation and Co-Creation”, *Mensch und Computer* 2023, 534–538, 2023. <https://doi.org/10.1145/3603555.3608572>
- [2] K. Khutsishvili, N. Pavicic, M. Combé, “The challenge of co-creation: How to connect technologies and communities in an ethical way”, in *The Leading Role of Smart Ethics in the Digital World*, Universidad de La Rioja, 2024.
- [3] M. D. Wilkinson, et al., “The FAIR Guiding Principles for scientific data management and stewardship”, *Scientific Data*, 3, 160018, 2016. <https://doi.org/10.1038/sdata.2016.18>

#### Please contact:

Rita Stampfl

University of Applied Sciences Burgenland, Austria  
[rita.stampfl@hochschule-burgenland.at](mailto:rita.stampfl@hochschule-burgenland.at)

## Services for Open Science Education and Skills Development: Beyond Technical Training

by Katharina Flicker (TU Wien, SBA), Ilire Hasani-Mavriqi (TU Graz), Dimitri Prandner (JKU)

***As Open Science becomes a structural requirement, developing the right skills goes beyond technical training and requires integrating disciplinary expertise with legal, communicative, and data-related competences.***

Over the last two decades, Open Science and open data have evolved from normative ideals into structural requirements for contemporary research systems. Policy frameworks increasingly mandate openness and promote infrastructures to generate public value from scientific work with open data [1]. However, competence development is often narrowly framed as technical training in research data management or FAIR compliance. This perspective has been criticised as insufficient [2], as sustainable Open Science education requires integrating disciplinary expertise with managerial, communicative, legal, and technical competences.

#### Infrastructure as a Precondition for Open Science Skills

Arguments emphasising the relevance of disciplinary, managerial, communicative and legal skills in Open Science and open data education are echoed in the findings from the recent quantitative RI:TRAIN PLUS [L1] survey study [2], which included data from 330 operators and managers of selected European research infrastructures (RI). Based on a comprehensive needs assessment across European RIs, the study shows that training demands for infrastructures fostering Open Science and open data are closely tied to the institutional maturity of infrastructures: if governance, data stewardship, repository services, and

policy frameworks are underdeveloped, skills development remains fragmented and inconsistent.

This suggests structural dependencies: Open Science education presupposes infrastructures that themselves operate according to FAIR principles and transparent governance standards. Researchers cannot be expected to implement FAIR data practices if repositories are unstable, metadata standards are unclear, or long-term preservation strategies are absent. Similarly, training in FAIR and open data stewardship is ineffective without institutional support structures that enable compliance in practice.

In this sense, open data education must follow infrastructural development. Services for Open Science and open data skills development must therefore be embedded within functioning ecosystems providing operational clarity, stable tools, and harmonised policies. Investments in training without parallel investments in infrastructural maturity risk producing symbolic compliance rather than sustainable transformation and vice versa.

### Beyond Technical Skills: Management, Communication, and Legal Competence

The data also indicate that Open Science competencies extend beyond technical data handling. As illustrated in Figure 1, participants repeatedly emphasised needs related to project management, stakeholder coordination, communication strategies, and legal clarity. These findings challenge a reductionist understanding of Open Science as a purely technical domain. It also provides deeper insight into the fact that overall actors in the field of Open Science require structural competences that enable them to bridge normative commitments and organisational realities [2, 3]. FAIR implementation, for instance, is not merely a metadata exercise but a governance question: Who defines and accepts standards? Who ensures compliance? Who bears responsibility for long-term sustainability? Likewise, legal questions regarding data protection, intellectual property, and licensing are central to enabling reuse and trust. Communication plays a similarly crucial role. Open Science presupposes dialogue across disciplines, institutions, and na-

tional contexts. It requires shared vocabularies, alignment of expectations, and the capacity to translate technical requirements into actionable guidance for diverse stakeholder groups. Without communicative infrastructures, even well-designed technical services remain underutilised. Thus, services for Open Science education must adopt a broadened competence model. In addition to data literacy and FAIR implementation, they must foster:

- governance and management skills for coordinating Open Science processes
- legal literacy regarding data protection, copyright, and licensing
- communication capacities to mediate between researchers, infrastructures, and policy actors.

Only by integrating these dimensions can Open Science move from the uncomfortable place between normative positions and expected policy-based regulations to operational practice.

### The Role of National Support Structures

National coordination structures play a decisive role in operationalising such integrated competence models. Positioned between European developments and national stakeholders, the EOSC Support Office Austria (SOA) [L2] supports Austrian institutions in aligning with the evolving EOSC Federation [L3], a European network of interoperable data repositories and services intended to advance Open Science. Services provided by EOSC SOA currently include communication support, structured information dissemination, and the coordination of stakeholder dialogues within the national Open Science community.

Beyond communication and stakeholder coordination, Austria has invested in complementary ministry-funded initiatives that strengthen the structural foundations of Open Science. The Shared RDM Services & Infrastructure project [L4], for example, develops interoperable and scalable research data management services across institutions, thereby creating the operational backbone required for effective training uptake. Likewise, the FAIR Data Austria initiative has contributed to

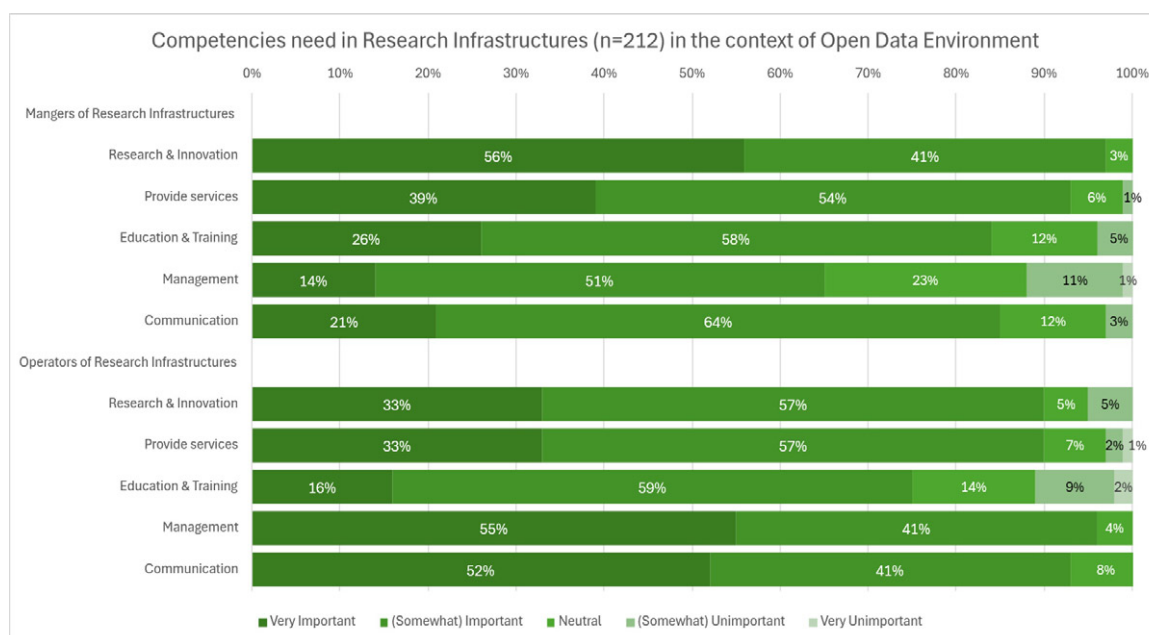


Figure 1: Distribution of competencies required to work in research infrastructures and to support Open Science. Results from the European RI:train plus survey 2021.

the professionalisation of data stewardship by establishing a national strategy and competence framework that outlines data stewardship models and associated training pathways [3].

Together, these initiatives illustrate that national support structures do not merely disseminate information but actively shape the infrastructural and organisational conditions under which Open Science skills can be sustainably developed. As also highlighted by the RI:TRAIN PLUS findings, competence development depends on infrastructural maturity and coherent governance frameworks and is not solely a matter of individual researcher training. National initiatives thus translate European-level ambitions into nationally embedded service portfolios, governance models, and professional profiles.

#### From Skills to Systems

Open Science education should therefore be conceptualised not as a discrete training activity but as part of a systemic transformation process. Technical FAIR skills remain essential. Yet without mature infrastructures, coherent governance, legal certainty, and communicative coordination, such skills cannot unfold their full potential.

Services for Open Science education must consequently operate on two levels: they must enhance individual competences while simultaneously strengthening institutional and infrastructural capacities. The data from RI:TRAIN PLUS indicates that these dimensions are interdependent. The example of the EOSC SOA, alongside initiatives such as Shared RDM Services & Infrastructure and FAIR Data Austria, shows how coordinated national support structures can address this interdependence in practice. If Open Science is to become the default mode of research rather than an additional administrative burden, education and infrastructure must evolve in sync. Only then can openness become operationally sustainable.

#### Links:

- [L1] <https://kwz.me/hIp>
- [L2] <https://eosc-austria.at/>
- [L3] <https://eosc.eu/building-the-eosc-federation>
- [L4] <https://forschung-daten.at/en/shared-rdm/>

#### References:

- [1] S. Reichmann, et al., “Adopting EOSC on the ground – Integrating (European) research infrastructure into national institutions,” *ABI Technik*, vol. 45, no. 4, pp. 432–442, 2025, doi: 10.1515/abitech-2025-0059.
- [2] D. Prandner and P. Sinner, “Identifying and updating training needs in European research infrastructures and core facilities,” *Zenodo*, 2022, doi: 10.5281/zenodo.7542429.
- [3] A. Bardel, et al., “Data stewardship – Austrian national strategy and alignment,” *Zeitschrift für Hochschulentwicklung (Journal for Higher Education Development)*, vol. 18, Sonderheft Forschung, pp. 65–88, 2023, doi: 10.21240/zfhe/SH-F/05.

#### Please contact:

Ilire Hasani-Mavriqi, TU Graz, Austria  
ilire-hasani-mavriqi@tugraz.at

Dimitri Prandner, JKU, Austria  
dimitri.prandner@jku.at

## Shaping the Czech National Infrastructure for Research Data through Community Consensus

by Martin Dvořák, Pavlína Špringerová and Matej Antol (Masaryk University)

*Four years ago, the Czech Republic laid out its plan for a National Data Infrastructure for FAIR (Findable, Accessible, Interoperable, Reusable) research data under the Czech chapter of the European Open Science Cloud initiative (EOSC CZ). After years of planning and development, the first comprehensive national survey of researcher practices delivers insight into its progress: although two-thirds of Czech researchers declare some familiarity with the Open Science principles in research data management, more than half still store research data primarily on portable devices, and a similar proportion have never heard of EOSC. These findings are directly influencing the initiative’s approach to working with the scientific community and its deployment of infrastructure, tools, and services.*

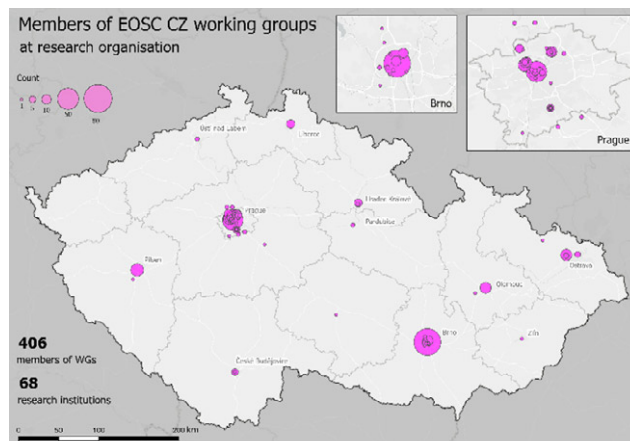


Figure 1: Members of the EOSC CZ Working Groups.

Since 2022, twelve working groups (WGs) have been established to identify the most pressing needs of Czech researchers regarding FAIR data management. More than 400 WG members drafted a comprehensive plan that encompasses hardware infrastructure, repository systems, data repositories, accompanying services and systemic support for data management training. The EOSC CZ initiative has received funding from the Ministry of Education, Youth and Sports (MEYS) through a series of closely interconnected projects. The whole effort is coordinated by Masaryk University and currently encompasses over 20 institutional partners. Its ambition is clear: to provide every Czech researcher with the necessary infrastructure, tools, and skills to manage data responsibly and in accordance with FAIR principles.

#### Build-up of the infrastructure

The national data infrastructure for FAIR research data is steadily taking shape. Its repository-as-a-service platform, including repo-systems CESNET Invenio, CLARIN-DSpace,

ASEP/ARL, and Islandora, is in pilot production and already provides approximately 10% of its 50 PB target storage capacity. It currently hosts a generalist repository [L1] and domain-specific repositories for biodiversity, molecular biophysics, and archaeological data. These repositories are further complemented with services for data processing, authentication and authorisation infrastructure, data management planning tools, metadata models and more. Interoperability is ensured by the newly introduced Czech Core Metadata Model (CCMM) [L2].

To build the human capacity needed alongside the technical infrastructure, the EOSC CZ Training Centre has trained more than 3,000 researchers, data stewards, curators, and enthusiasts, who, in turn, bring new sets of requirements for the infrastructure and services. Alongside the EOSC CZ Working Groups, a national data steward community has been established and has grown to over 100 professionals [L3] actively improving data management processes and standards within research institutions. Moreover, a new research assessment methodology provided by MEYS now recognises datasets as a standalone result type, providing a systemic incentive for awareness of the importance of research data.

### Reflecting on the current progress

While the infrastructure is being built, the EOSC CZ initiative conducted the most comprehensive assessment of Open Science and FAIR data practices in the Czech Republic to date, surveying 1,121 principal investigators, with a 29.5% response rate [2]. The results reveal a persistent gap between awareness and practice, reflecting slow adaptation of research workflows to FAIR principles (Figure 2).

About 58% of researchers report familiarity with FAIR data principles, yet two-thirds (66%) still store research data primarily on personal computers and 53% use portable devices (USB drives, CD/DVD, or external hard drives), exposing them to risks of loss, damage, or theft. National e-infrastructure remains underutilised, with only 14.5% adopting it during their projects. Repository adoption varies sharply by discipline: 64% of life science scientists upload data to repositories, compared to 39% in the humanities. Data Management Plans split the community – 44% see them as useful, 54% as a bureaucratic burden – though researchers with access to data stewards are significantly more likely to embrace FAIR practices. Notably, 62% of researchers are unfamiliar with EOSC CZ, the very initiative that is building the infrastructure for them.

The perception of bureaucracy also extends beyond data management planning. When asked what prevents re-

## Awareness about Open Science

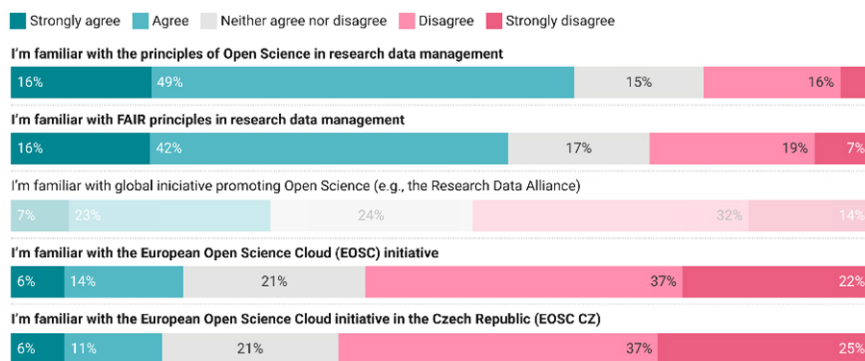


Figure 2: Awareness about Open Science.

searchers from sharing their data, researchers primarily point to administrative and legal concerns – copyright and licensing issues top the list of obstacles, flagged by 63% of respondents. Over 40% of respondents simply do not know which repository to use, directly mirroring the survey's finding that 62% are unfamiliar with EOSC CZ itself. Fears of misinterpretation and misuse of data (around 50% and 40%, respectively) suggest that reluctance to share is not merely administrative but rooted in a broader lack of trust and guidance (Figure 3).

### The imminent future

The survey findings make clear that building technical infrastructure is not enough. Researchers' reluctance to share data stems not only from practical gaps, not knowing which repository to use or how to structure data, but from deeper concerns about trust, misinterpretation, and administrative burden. Shifting this mindset is as important as deploying the tools themselves, and the infrastructure must be designed to actively address these concerns rather than simply meet technical standards.

The community has been established, and the infrastructure is entering its operational phase. Its real value, however, will materialise only if the barriers and needs identified through the surveys are systematically addressed, and the heterogeneity of

## What kind of obstacles do you see in sharing the research data?

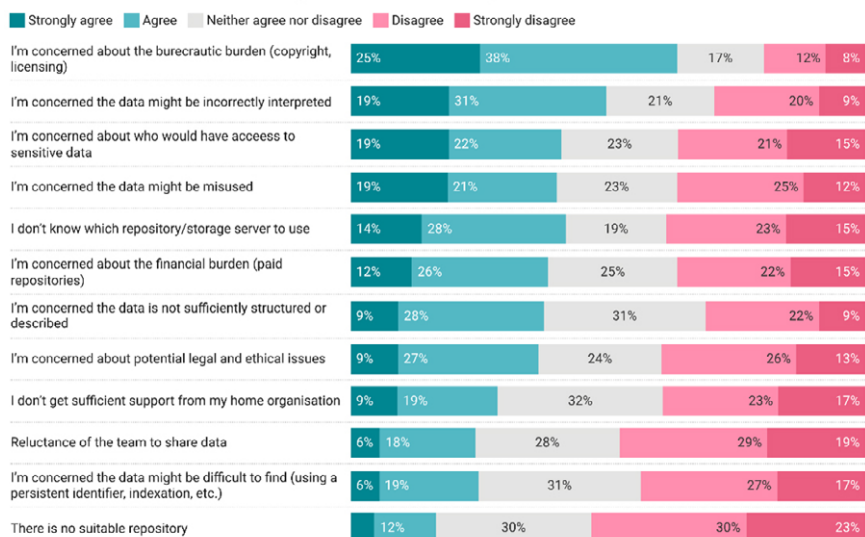


Figure 3: Perceived obstacles in sharing the research data.

scientific domains and their specific requirements is fully acknowledged. In the end, building a national infrastructure for research data requires continuous alignment between researchers' evolving needs and the initiative's design, services, and priorities.

#### Links:

[L1] <https://datarepo.eosc.cz>

[L2] <https://www.ccm.cz/en/>

[L3] <https://arcg.is/0HGKPr3>

#### References:

[1] M. Antol et al., "EOSC CZ: Towards the development of Czech national ecosystem for FAIR research data," 2024, <https://arxiv.org/abs/2402.13343>.

[2] M. Vávra et al., "Průzkum správy výzkumných dat v České republice 2025," CSDA, 2026. <https://doi.org/10.14473/CSDA/OJTMJF>.

#### Please contact:

Martin Dvorak, Masaryk University, Czech Republic  
[dvorak@ics.muni.cz](mailto:dvorak@ics.muni.cz)

## Creative Commons Licences in the Age of AI: Challenges and Opportunities

by Daniel Spichtinger (Ludwig Boltzmann Gesellschaft / University of Vienna/independent researcher)

***Creative Commons licences underpin the reusability of research data and are a key enabler of Open Science, yet the rise of large-scale AI training is testing their limits. This article, based on and updating a recent book chapter [1][LC1.1][DS1.2], examines the legal landscape in the European Union and the United States.***

Licensing plays a central role in ensuring reusability: without clear, machine-readable licenses, research outputs risk becoming technically interoperable but legally unusable, undermining reproducibility and trust, which are at the core of Open Science. [DS2.1] In the context of publicly funded research, Creative Commons (CC) licenses provide a standardised framework that enables predictable and legally secure reuse.



Figure 1: CC Stickers, Illustration by Kristina Alexanderson on Flickr.

Therefore, research funders such as the European Commission require CC licences (e.g. in Horizon Europe) as a funding condition and consequently many researchers provide CC-licences for their research outputs (most commonly CC-BY).

#### Litigate or legislate – US and EU perspectives

CC licences were designed before the recent wave of commercial AI systems using large language models emerged. The question that has thus informed my work is: can CC-licensed research output be legally used by AI systems to train them? This is further complicated by different legal regimes for the use of such data in AI training in the United States and the European Union. These developments raise important questions for Open Science, as legal uncertainty around data use may affect the willingness of researchers to share data openly.

In the US, the fair use doctrine may permit the use of copyrighted material for AI training if the use is deemed “transformative.” Given that Creative Commons licences are a part of the copyright system this also applies to them. Invoking the fair use doctrine, companies could therefore ignore CC-licence attributions, such as non-commercial restrictions. In 2019, IBM researchers used over 1 million CC-licensed Flickr photos to train a facial recognition system without notifying photographers [2].

However, whether the use of data by AI companies really constitutes fair use is legally contested, with several court cases still pending. It appears, however, that companies having invoked fair use prefer to settle outside court rather than waiting for a definitive decision by the court. For example, the New York Times has signed a licensing deal with Amazon for AI use of its content.

The EU takes a more structured approach. The 2019 Directive on Copyright in the Digital Single Market[LC3.1][DS3.2] [3] introduced text and data mining (TDM) exceptions: Article 3 permits TDM for non-commercial scientific research, while Article 4 allows commercial TDM provided that rights holders have not opted out. Both of these options could therefore also be used to circumvent CC licencing provisions. The EU’s AI Act[LC4.1][DS4.2],[L1] which entered into force in August 2024, builds on this framework. Article 53(1)(c) requires providers of general-purpose AI models to identify and respect opt-out reservations, even if they are based outside Europe.

From August 2025, AI providers must have copyright compliance policies in place. A General-Purpose AI Code of Practice[LC5.1] [L2], published in July 2025, offers further guidance, including standardised machine-readable opt-out mechanisms. However, The EU’s AI Office has stated that during its first year (until August 2026) it will not consider providers to have broken their commitments and will not adopt measures against them if they do not fully implement all commitments immediately after signing the Code [L3].

While these frameworks aim to balance innovation and rights, they also introduce complexity that may challenge the straightforward reuse of openly licensed research outputs.

#### Can AI-Generated Research Output Be CC-Licensed?

A related question concerns whether AI-generated outputs (inter alia generated by researchers) can themselves be copy-

righted and thus CC-licensed. In 2023, the US Copyright Office ruled on *Zarya of the Dawn*, a graphic novel with AI-generated images: while the human-authored text was copy-rightable, the AI-generated images were not, due to insufficient human creative involvement. Similarly, under EU law, originality requires a work to reflect the author's own intellectual creation. Outputs generated without sufficient human input therefore fall outside copyright protection and effectively enter the public domain; therefore they cannot be CC-licensed. Researchers using AI tools should therefore carefully document their creative contributions to maintain any claim to authorship. This has implications for Open Science, as the status of AI-generated outputs affects their reuse, attribution, and integration into open research workflows.

#### Emerging Responses and Outlook

Several initiatives are emerging to address creative commons licencing in the age of AI. Most notably, Creative Commons itself has developed “preference signals” – machine-readable metadata tags indicating whether CC-licensed content may be used for AI training. The EU AI Act's extraterritorial reach could turn these signals into globally binding indicators, much as the GDPR has influenced many personal data protection standards around the world.

Meanwhile, the US remains without comprehensive federal AI legislation. Executive Order 14,179 (January 2025) and the America's AI Action Plan (July 2025) emphasise deregulation and competitiveness, while proposals requiring copyright disclosure in AI training datasets remain stalled in Congress.

CC licences remain important for ensuring that research outputs [LC6.1][DS6.2] are not only technically interoperable but also legally reusable, a cornerstone of Open Science. Adapting them for AI will be key to preserving the culture of Open Science while safeguarding creators' rights.

#### Links:

[L1] <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

[L2] <https://kwz.me/hIk>

[L3] <https://kwz.me/hIn>

#### References:

- [1] D. Spichtinger, “Fit for Purpose? Creative Commons Licensing for Research Data in the Age of Artificial Intelligence,” in *Data Quality Matters*. London, U.K.: IntechOpen, 2025. doi: 10.5772/intechopen.1013402.
- [2] R. Merkle, “Use and fair use: Statement on shared images in facial recognition AI,” *Creative Commons Blog*, Mar. 13, 2019. [Online]. Available: <https://creativecommons.org/2019/03/13/statement-on-shared-images-in-facial-recognition-ai/>
- [3] European Parliament and Council of the European Union, “Directive (EU) 2019/790 of 17 April 2019 on copyright and related rights in the Digital Single Market,” *Official Journal of the European Union*, L 130, pp. 92–125, May 17, 2019. [Online]. Available: <https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng>

#### Please contact:

Daniel Spichtinger, Ludwig Boltzmann Gesellschaft/  
University of Vienna/Independent Researcher, Austria  
[dspichtinger@outlook.com](mailto:dspichtinger@outlook.com)

## LICORICE: Deploying Privacy-Enhancing Technologies for Europe's Digital Sovereignty

by Diana E. Jimenez-Bejarano (Trialog), Stephan Krenn (AIT Austrian Institute of Technology), Antonio Kung (Trialog), and Angel Palomares Perez (Advanced Computing SL)

*Privacy-Enhancing Technologies (PETs) have moved from academic prototypes to relevant building blocks of Europe's digital policy agenda. Developments such as GDPR, the NIS2 Directive, the EU AI Act and the European Digital Identity framework show a clear direction: using or analysing sensitive data must not weaken confidentiality, privacy, data sovereignty, or compliance. At the same time, modern AI systems increasingly rely on cross-organisational collaboration. This dynamic creates a structural tension between data utility and privacy protection.*

The Horizon Europe project LICORICE [L1] addresses the tension between data utility and privacy protection by bringing advanced cryptographic technologies into concrete, policy-relevant settings. The project develops interoperable tools for privacy-preserving identity management and for secure computation on sensitive data.

#### From Fragmented Research to Deployable Tools

In recent years, cryptographic research has produced powerful methods such as secure multi-party computation, homomorphic encryption, differential privacy and zero-knowledge proofs. Still, many of these approaches remain rather fragmented and strongly focused on theoretical questions. Integration into existing infrastructures is often complex, and usability or regulatory alignment is often underdeveloped.

LICORICE follows a more pragmatic approach. Instead of refining isolated primitives, the project combines mature and emerging PETs into coherent and deployable tools. This supports compliance with evolving European frameworks, including EUDI wallets, sector-specific data spaces and cybersecurity obligations under NIS2.

Two complementary technology stacks are developed:

- a privacy-oriented identity management toolset aligned with Self-Sovereign Identity concepts and European Digital Identity initiatives
- a secure computation toolset including secure federated learning, multi-party computation and secure neural network inference.

Together, these components form a practical foundation for trusted data ecosystems in regulated domains. The technologies are validated through two pilot implementations: one in healthcare and one in cybersecurity threat intelligence sharing.

#### Pilot 1: Privacy-Preserving AI for Healthcare

Healthcare data is among the most sensitive categories of personal data. Electronic Health Records, imaging data and infor-

mation from wearable devices are essential for AI-supported medicine. At the same time, regulatory frameworks such as GDPR, HIPAA and the EU AI Act may restrict data sharing between hospitals and institutions. As a consequence, valuable datasets often remain siloed, limiting robust machine learning development.

LICORICE tackles this challenge in the context of chronic respiratory diseases, notably Chronic Obstructive Pulmonary Disease (COPD) and asthma, conditions that affect hundreds of millions of people worldwide and place a significant burden on healthcare systems. Earlier detection of exacerbations through AI-based prediction can reduce hospitalisations and costs, especially in ageing societies.

The pilot combines:

- self-sovereign identity mechanisms for strong and privacy-preserving authentication of patients and healthcare professionals, including cryptographically protected biometrics
- federated learning with enhanced privacy guarantees, allowing hospitals to jointly train models without transferring raw patient data
- secure aggregation based on multi-party computation, so that only aggregated model updates are visible.
- secure neural network inference [1], protecting both patient input data and proprietary models during prediction.

The scenario integrates representative hospital records and data streams from wearable devices measuring parameters such as ECG or oxygen saturation. Distributed machine learning across institutions becomes feasible while maintaining confidentiality and data sovereignty. The integration is illustrated in Figure 1.

From a policy angle, this pilot gives a concrete example of how a European Health Data Space could enable cross-border analytics without centralising sensitive information. It shows, in practical terms, that AI innovation and fundamental rights protection need not contradict each other.

### Pilot 2: Privacy-Preserving Cyber Threat Intelligence Sharing

The second pilot focuses on cybersecurity, where rapid information exchange is crucial to counter increasingly sophisticated attacks. AI-based cybersecurity assistants analyse large volumes of logs and threat intelligence to detect anomalies and predict

vulnerabilities. However, these systems often depend on access to sensitive operational and, in some cases, personal data.

In practice, automated processing is often restricted once confidential information is present. This leads to manual review and slower incident response. At the same time, effective cyber defence depends on structured information sharing across organisational and national borders, which again raises privacy and compliance concerns.

LICORICE develops a privacy-preserving CTI assistant that addresses both issues. The system includes:

- automated anonymisation of queries sent to large language models provided as Machine Learning as a Service [2], enabling the safe use of external AI services
- differential privacy and multi-party computation techniques for secure data sharing across organisations
- integration with the MISP threat intelligence platform, allowing collaborative analysis without exposing sensitive underlying data.

By embedding PETs directly into threat intelligence workflows, LICORICE demonstrates that AI-driven security tools can remain compliant with data protection requirements while keeping operational efficiency.

This is particularly relevant in view of NIS2, which strengthens cybersecurity obligations across critical sectors and encourages structured information sharing. The pilot provides technical mechanisms to implement such sharing in a privacy-preserving way.

### Towards Trustworthy Digital Infrastructures

Europe’s regulatory landscape is often perceived as a constraint on innovation. LICORICE instead treats regulation as a design requirement. By embedding compliance, accountability and privacy into system architectures from the outset, the project aligns technological capability with European values, proving that privacy and utility are not mutually exclusive.

In this sense, the project contributes to digital sovereignty in a concrete way: enabling advanced analytics and cross-border collaboration without losing control over sensitive data. The developed toolsets aim to be interoperable and practically usable within emerging European data spaces and identity frameworks.



Figure 1: Secure framework for access and storage and storage of data in the LICORICE health pilot [3].

As PETs continue to mature, large-scale deployment will depend less on new theoretical results and more on integration, usability and regulatory compatibility. LICORICE positions itself at this interface, translating cryptographic research into operational trust.

#### Links:

[L1] <https://www.licorice-horizon.eu/>

#### References:

- [1] K. Batool, S. Anwar, and Z. Á. Mann, “SecFePAS: Secure facial-expression-based pain assessment with deep learning at the edge,” in Proc. SEC 2024, 2024, pp. 417–424.
- [2] V. Prodomo, R. Gonzalez, and M. Gramaglia, “SCIPER: Secure collaborative inference via privacy-enhancing regularization,” IEEE Trans. Privacy, vol. 1, pp. 57–68, 2024.
- [3] LICORICE Consortium, “D3.2 Pilot planning, requirements and LICORICE toolset definition,” Project Deliverable, 2025.

#### Please contact:

Diana E. Jimenez-Bejarano, Trialog, Paris, France  
[diana.jimenez@trialog.com](mailto:diana.jimenez@trialog.com)

## The EOSC EU Node: An Evolving Gateway to Open Science in Europe

by Alane Brunschweiler, Maja Dolinar and Fay Meimaraki (OpenAIRE AMKE)

***One year after its launch, the EOSC EU Node is emerging as a key European gateway for Open Science, enabling researchers to access, share, and collaborate across borders.***

The European research landscape is undergoing a structural transformation. Open Science is no longer a policy aspiration, but an operational requirement: research outputs must be findable, accessible, interoperable and reusable (FAIR) across institutional and national boundaries. Delivering this transformation requires more than just advocacy, it demands shared digital infrastructure that works in practice. The European Open Science Cloud (EOSC) EU Node represents a major milestone in this effort.

Launched officially at the EOSC Symposium in Berlin in October 2024, the EOSC EU Node became the first operational node of the EOSC Federation. It provides an EU-level access point where researchers can discover and use scientific research outputs, computing resources, and collaborative environments in one place. Developed and operated by the European Commission through a procurement framework involving multiple European service providers, the EOSC EU Node is designed as both a functional research environment and a reference model for how a federated European research infrastructure can operate.

For the first time, researchers across Europe can access a unified space where scientific research outputs, computing services and collaborative workspaces can be discovered and used through a single access point. Tens of millions of research outputs are discoverable in one place, alongside interactive services such as virtual machines, data storage, file synchronization, notebook environments and collaborative workspaces, all available free at the point of use (subject to the virtual credits mechanism).

Since its launch, the EOSC EU Node has continued to evolve in response to feedback from its growing user community. This user-driven approach has helped shape the platform into a more intuitive and practical environment aligned with real research needs.

One of the most visible results of this evolution is the redesign of the User Space and overall user experience. A unified dashboard now provides users with a consolidated overview of projects, storage allocation, computing sessions, and shared workspaces. This reduces onboarding friction and allows research teams to move more quickly from discovery to active collaboration.

Alongside these usability improvements, the EOSC EU Node has updated its User Access Policy to support a broader range of research workflows. Credit allocations, virtual units you can use to access services on the EOSC EU Node, have nearly doubled for all user categories. This allows individual contributors and research teams to plan more ambitious work, run longer computing tasks, and experiment with larger datasets without encountering resource limits too early.

Another important development has been the introduction of credit grants for EU-funded projects. These grants of up to 40,000 credits provide research groups with enhanced access to the services of the EOSC EU Node, making it easier to integrate Open Science tools and practices from the start of a project rather than later in its lifecycle.

Users are supported through extensive documentation and a dedicated learning platform that brings together structured courses, practical tutorials, and optional learning pathways tailored to different user needs and roles. In addition, targeted guidance resources provide step-by-step support on how to integrate services into everyday research workflows. Together, these materials ensure that researchers can confidently adopt and apply the tools of the platform, regardless of their level of technical expertise.

This user-driven and community-responsive approach reflects a broader ambition: to build a shared infrastructure that evolves alongside its community. Rather than imposing fixed technological solutions, the EOSC EU Node continues to adapt based on how researchers work, ensuring that services remain relevant and accessible across disciplines.

Beyond its direct benefits to researchers, the EOSC EU Node plays an important systemic role within the wider EOSC Federation. As the first operational European node, it serves as a reference model for future national and thematic nodes that

will gradually join the Federation. Together, these interconnected nodes will form a distributed ecosystem built on shared standards and principles, enabling seamless access to research resources across Europe.

Looking ahead, future developments will further strengthen the position of the EOSC EU Node within the European research landscape. These include expanded integration of trusted third-party services and resources, continued enhancement of user experience, and stronger engagement with research communities across Member States. The aim is to ensure that the EOSC EU Node remains a flexible and reliable environment that supports new ways of working with data, software, and collaborative platforms. As additional national and thematic nodes join the EOSC Federation, coordination and governance will become increasingly important. The EOSC EU Node is expected to continue playing a stabilising role, demonstrating operational practices, testing interoperability mechanisms, and supporting adoption at scale.

Since October 2024, the EOSC EU Node has moved from an initial concept to a functioning infrastructure supporting researchers across Europe. By lowering practical barriers to accessing and sharing research outputs, it contributes to making Open Science not only a principle but a daily working reality. As the EOSC Federation continues to grow, the EOSC EU Node is positioned to play a key role in supporting collaboration, discovery, and innovation across the European research landscape.

OpenAIRE AMKE is the member of the consortium contracted for the EOSC EU Node services of the European Commission. This article reflects the views only of the authors and not of the European Commission.

**Link:**

To explore the EOSC EU Node and its services, visit:  
<https://open-science-cloud.ec.europa.eu/>

**Please contact:**

Fay Meimaraki, OpenAIRE AMKE  
[fay.meimaraki@openaire.eu](mailto:fay.meimaraki@openaire.eu)

## The EOSC Core Innovation Sandbox: A Launchpad for Europe's Open Science Federation

by Federico Drago (EGI Foundation) and Nicola Fiore (EGI Foundation)

***Becoming part of the EOSC Federation is a significant step for any organisation involved in research. The new EOSC Core Innovation Sandbox provides a safe, pre-production environment where providers can test their integration, validate interoperability, and gain confidence before joining the live federation, turning the complexity of Open Science into a manageable, step-by-step process.***

The European Open Science Cloud (EOSC) is transitioning from a centralised platform to a federation of autonomous, interoperable Nodes. For national data centres, thematic research infrastructures, and e-infrastructures, this shift presents a major opportunity. However, it also introduces a challenge: how can an organisation rigorously test its services against EOSC standards and ensure interoperability without risking disruption to production systems?

The EOSC Beyond project [L2] directly addresses this issue with the EOSC Core Innovation Sandbox [L1, Figure 1]. Launched operationally in spring 2025, the Sandbox is a comprehensive, pre-production testbed that mirrors the production EOSC environment. It acts as a "Federator Test Node," providing a complete suite of Core Federating Capabilities that any aspiring EOSC Node can use to validate its integration in a safe, controlled, and governed space.

### A Realistic Environment for Real-World Testing

Unlike static documentation or isolated test harnesses, the Sandbox provides a living ecosystem [1]. It hosts three distinct environments: Development, Integration, and Pre-production, allowing providers to progress from individual component testing to full end-to-end validation. Critically, this infrastructure is distributed; each Core service provider deploys their own instances on their own premises, exactly as they would in the production federation. This approach ensures that tests conducted in the Sandbox accurately reflect the conditions of the live EOSC.

A distinctive feature of the Sandbox is its operational role as a "Federator Test Node" within the EOSC architecture. In the production federation, Federator Nodes provide the shared capabilities (such as AAI, Monitoring, and the Helpdesk) that enable autonomous Nodes to interoperate.

The Innovation Sandbox replicates this function in a pre-production setting, offering a fully functional instance of these Federating Capabilities. This allows candidate Nodes to experience, first-hand, how their local services would connect to and interact with the federation's core infrastructure. By testing against a live Federator Node in a safe environment, providers can validate compliance with interoperability guide-

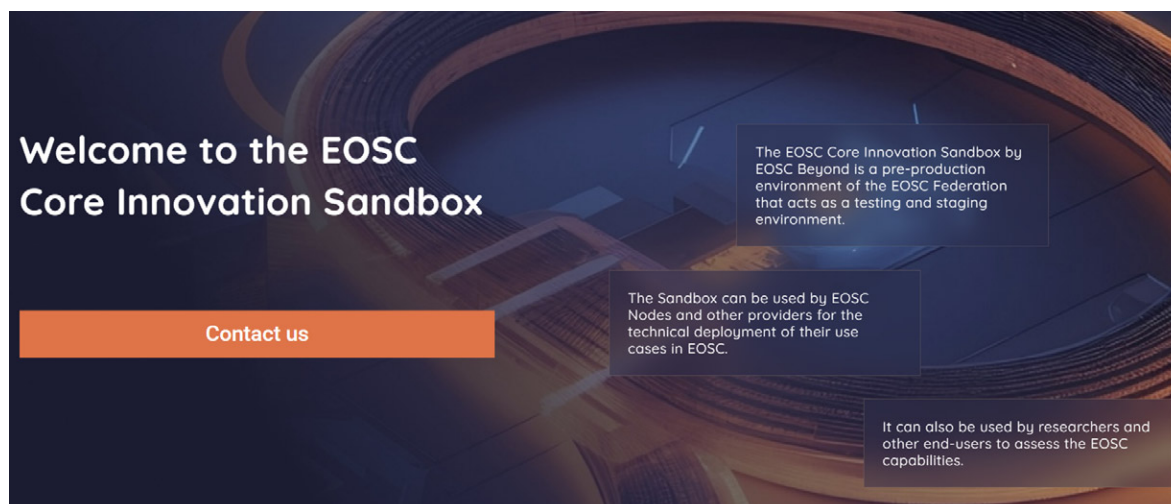


Figure 1: EOSC Core Innovation Sandbox landing page.

lines, refine their integration strategies, and resolve technical issues before they ever impact production systems. This approach transforms federation from an abstract concept into a tangible, testable process.

#### Lowering Barriers with Core Capabilities

The Sandbox is equipped with a full stack of EOSC Core services, enabling providers to tackle the most common and complex integration challenges head-on. Key capabilities available for validation include:

- **Federated AAI:** Providers can integrate their existing identity management systems with the EOSC Authentication and Authorisation Infrastructure, enabling Single Sign-On and supporting advanced security features like Multi-Factor Authentication.
- **Federated Helpdesk:** The Sandbox allows organisations to test the federation of their local ticketing systems with the pan-EOSC Helpdesk, ensuring that users receive coordinated support regardless of which Node they access.
- **Federated Monitoring:** Providers can validate how their services report operational status and availability to the central EOSC Monitoring framework, a prerequisite for building trust and transparency across the federation.
- **Resource Discovery & Onboarding:** Through enhanced registries and the EOSC Knowledge Graph, the Sandbox simplifies the process of making datasets, software, and services discoverable and machine-actionable.

These capabilities have already been successfully validated by a network of EOSC Pilot Nodes [3], including CESSDA, LifeWatch ERIC, NFDI, and e-INFRA CZ. Their successes (from enabling cross-domain data fusion to orchestrating "compute close to data" workflows) demonstrate that the Sandbox is not a theoretical exercise, but a proven, practical tool for de-risking federation.

#### An Open Invitation to the Community

The EOSC Core Innovation Sandbox is now open to external organisations. The process is intentionally lightweight: interested providers submit a brief use case via the project contact form, describing their community, technical requirements, and integration goals. Selected participants receive access credentials, onboarding guidance, and dedicated technical support from EOSC Beyond experts.

For service providers, this is a chance to validate new offerings. For research communities, it is an opportunity to test domain-specific workflows. For infrastructure operators, it is a direct pathway to becoming a functioning EOSC Node.

The success of the EOSC Federation depends not only on its technical design, but on the ease with which diverse organisations can join it [2]. The EOSC Core Innovation Sandbox transforms this complex process into a clear, supported, and low-risk journey. By providing a realistic testing ground and a complete suite of federating capabilities, it enables the next generation of EOSC Nodes to move from aspiration to operation, accelerating the transition toward a truly open and collaborative European research landscape.

#### Links:

[L1] <https://sandbox.eosc-beyond.eu/>

[L2] <https://eosc-beyond.eu/>

#### References:

[1] EOSC Beyond D5.4: "EOSC Core Innovation Sandbox" <https://doi.org/10.5281/zenodo.17292990>

[2] EOSC Federation Handbook. <https://doi.org/10.5281/zenodo.14999577>

[3] EOSC Beyond D15.2: "First report of the technical integration of the EOSC Beyond Pilot Nodes". <https://doi.org/10.5281/zenodo.17341191>

#### Please contact:

Nicola Fiore, EGI Foundation, Italy  
[nicola.fiore@egi.eu](mailto:nicola.fiore@egi.eu)

Federico Drago, EGI Foundation, Italy  
[federico.drago@egi.eu](mailto:federico.drago@egi.eu)

# EOSC Data Commons: Building Europe's Next-Generation Research Data Infrastructure

by Enol Fernández, Ilaria Fava and Xavier Salazar (EGI Foundation)

**EOSC Data Commons is a new European initiative to build the next generation of federated research data infrastructure. Researchers often know the data they need exists but cannot find it or access it easily. EOSC Data Commons tackles this fragmentation head-on by making cross-border discovery and reuse far more practical.**

EOSC Data Commons [L1] is a recently launched project that aims to shape the next generation of research data infrastructure in Europe. The project contributes to the European Open Science Cloud (EOSC) [1] by fostering seamless access to high-quality, interoperable research data and services. In practice, this means helping researchers in different fields actually work on the same data without spending weeks resolving format, access, or compatibility issues.

Coordinated by the EGI Foundation, EOSC Data Commons brings together a consortium of European infrastructure and renowned data repositories, virtual research environment providers and data analysis workflow providers [L2]. It links national and thematic infrastructures across Europe, creating a federated system that works across borders but respects how individual providers operate.

## Why EOSC Data Commons is needed

Research data volumes have grown dramatically over the last decade, driven by advances in digital technologies, large-scale instruments, simulations, and data-intensive methods such as artificial intelligence. While this data explosion holds enormous

potential, much of the resulting data remains siloed, heterogeneous, and difficult to discover, access, or reuse. As a consequence, valuable scientific insights often remain untapped, and reproducibility and cross-disciplinary collaboration are hindered.

The project addresses these challenges by creating a trusted, distributed environment that brings together data, tools, and compute resources across disciplines and national boundaries. By reducing fragmentation and improving interoperability, the project aims to make research data more FAIR [2] and support the full research data lifecycle, from deposition through long-term preservation and reuse.

## What EOSC Data Commons is building

At the core of EOSC Data Commons is the development of two complementary services.

- EOSC Matchmaker [L3] enables researchers to discover datasets, tools, and services across multiple scientific domains. This discovery layer is combined with a catalogue of analytics tools and execution services, allowing tools to be deployed close to the data.
- EOSC Data Player [L4] complements this by addressing interoperability and integration of execution platforms and data providers. It provides harmonised APIs, shared metadata specifications, and common mechanisms for authentication, authorisation, and provenance tracking. In practical terms, this means a researcher can combine data and tools from different providers without manually reconciling metadata formats or authentication systems, while still keeping full traceability.

The idea is to let researchers move from discovery to analysis and preservation without switching between disconnected systems. By bringing computation to the data, EOSC Data Commons services reduce data movement, improve efficiency, and enable scalable analysis across distributed infrastructures.

## Technical Approach

EOSC Data Commons employs a range of advanced technical approaches to realise its vision. These include distributed ar-

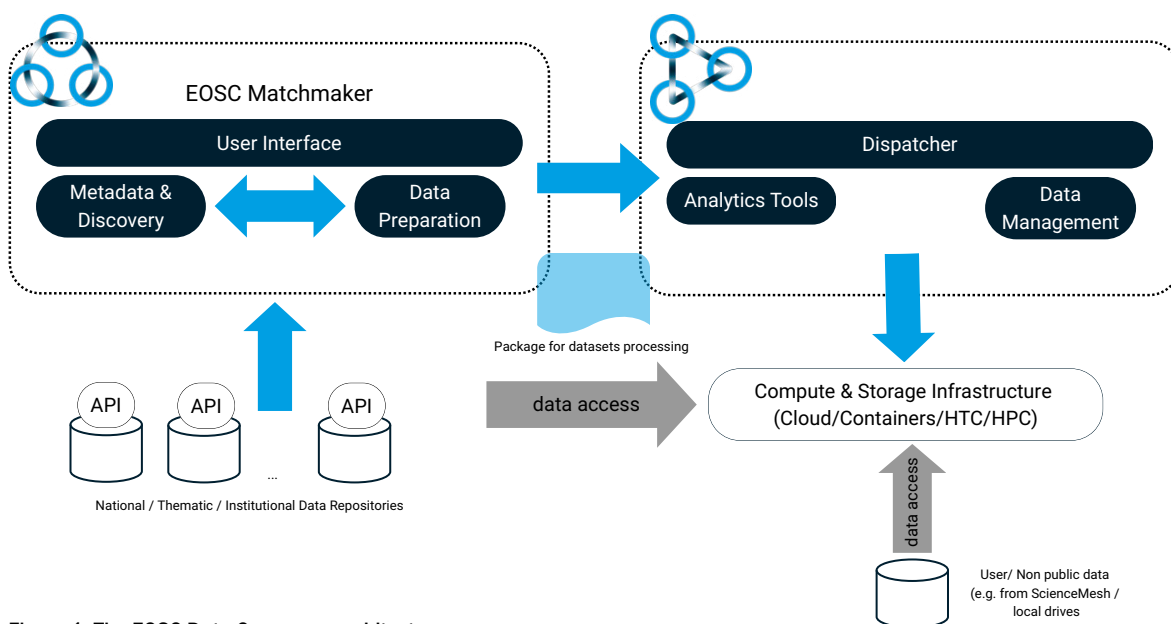


Figure 1: The EOSC Data Commons architecture.

chitectures that respect the autonomy of data providers while enabling cross-infrastructure integration, the use of semantic technologies and knowledge graphs for rich metadata representation and discovery; and standardised interfaces to support interoperability. The project adopts community standards such as RO-Crate [3] for packaging research artefacts and their metadata in a machine-actionable, interoperable manner.

The project also explores machine-assisted methods for search and analysis, particularly where datasets are too large or complex for manual exploration. EOSC Data Commons aims to help researchers extract insights more efficiently from large, complex datasets while improving reproducibility and transparency of research results. Figure 1 illustrates the architecture of EOSC Data Commons.

#### Orientation and Validation Through Use Cases

EOSC Data Commons is strongly user-driven. The project is validated through real-world use cases [L5] spanning a wide range of scientific domains, including life sciences, social sciences, environmental research, physics, and beyond. These use cases ensure that the developed services address concrete research needs and demonstrate cross-disciplinary relevance. By working closely with research communities, data repository managers, and service providers, the project aligns technical development with practical requirements and established research workflows.

#### Who Can Benefit and Participate

EOSC Data Commons is designed to benefit and engage a broad range of stakeholders. This includes researchers seeking interoperable access to data and compute resources; data repository managers and infrastructure providers interested in joining a federated European ecosystem and expanding its outreach to new communities beyond existing boundaries; tool developers and analytics service providers looking for integration opportunities; and research organisations and policy-makers committed to Open Science, data reuse, and FAIR principles.

The project is actively offering opportunities for participation and collaboration, strengthening the EOSC ecosystem.[L6].

#### Timeline and Future Activities

Started in April 2025, EOSC Data Commons is an ongoing effort whose results will evolve over the course of the project. Future activities include further integration of data repositories and services, expansion of supported use cases, and continued refinement of the EOSC Matchmaker and EOSC Data Player services based on user feedback.

In the longer term, EOSC Data Commons aims to provide a sustainable foundation for Europe's open research data landscape, supporting innovation, collaboration, and scientific excellence well beyond the project's lifetime.

#### Links:

- [L1] <https://www.eosc-data-commons.eu/>
- [L2] <https://www.eosc-data-commons.eu/about>
- [L3] <https://kwz.me/hIN>
- [L4] <https://kwz.me/hIe>
- [L5] <https://www.eosc-data-commons.eu/use-cases>

[L6] <https://www.eosc-data-commons.eu/open-call>

#### References:

- [1] Horizon Europe Co-programmed Partnership for the European Open Science Cloud (EOSC), “Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC),” ver. 1.3, Zenodo, 2024. doi: 10.5281/zenodo.17582648.
- [2] M. Wilkinson, et al., “The FAIR Guiding Principles for scientific data management and stewardship,” *Sci. Data*, vol. 3, p. 160018, 2016, doi: 10.1038/sdata.2016.18.
- [3] S. Soiland-Reyes, et al., “Packaging research artefacts with RO-Crate,” *Data Science*, vol. 5, no. 2, pp. 97–138, 2022, doi: 10.3233/DS-210053.

#### Please contact:

Enol Fernández, EGI Foundation, The Netherlands  
[enol.fernandez@egi.eu](mailto:enol.fernandez@egi.eu)

## ENVRI-Hub: A Science Gateway for Open Environmental Research

by Federico Drago (EGI Foundation), Delphine Dobler (Euro-ARGO ERIC) and Ulrich Bundke (Forschungszentrum Jülich / IAGOS AISBL)

***The new ENVRI-Hub platform, developed by the ENVRI-Hub NEXT project, serves as a central science gateway for Europe's environmental research. It demonstrates how interoperable Virtual Research Environments can accelerate collaborative Open Science and make FAIR data actionable for tackling global challenges.***

Environmental research is key to addressing planetary challenges, yet data and tools are often fragmented across disciplines and infrastructures. This inefficiency not only drains scientific time and resources but also erodes the core Open Science principles of transparency, collaboration, and reproducibility. The recently upgraded ENVRI-Hub platform [L1] directly addresses this by serving as an integrated science gateway and offering a Virtual Research Environment (VRE) interface for Europe's Environmental Research Infrastructures (ENVRIs).

Developed over the past two years by the Horizon Europe ENVRI-Hub NEXT project [L2], this platform is the result of extensive collaboration between leading research institutions, e-infrastructures, and ESFRI Landmarks across Europe. Its core aim is to operationalise Open Science principles by providing a unified, FAIR-compliant point of access.

The ENVRI-Hub enables users to search, access, and process datasets and services from across the ENVRI landscape. Users can be assisted by an AI Environmental Expert, which allows them to formulate their request and receive answers, centred on the ENVRI-Hub and Environmental RIs assets, in a natural language [1,2]. A core focus of the current development is en-

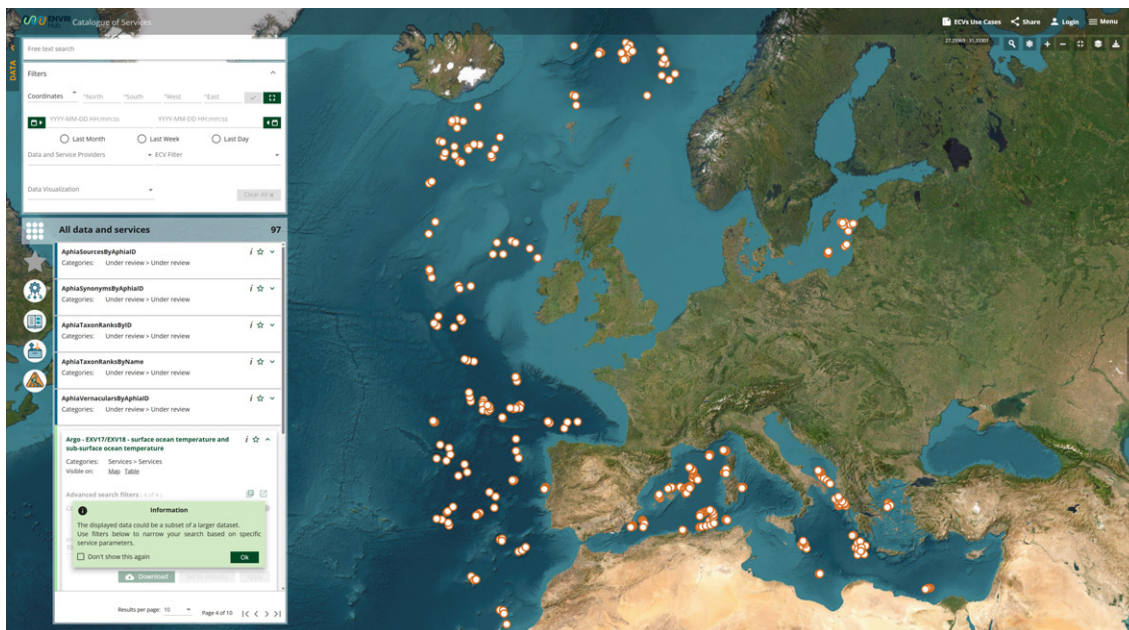


Figure 1: ENVRI Catalogue of Services providing unified access to environmental data and analytical tools across research domains.

ensuring that Essential Climate Variables (ECVs), critical indicators for climate research, are easily accessible and processable from all contributing ENVRI sources, supported by dedicated functionalities and workflows.

#### How the Hub Fosters Open Science Practices

The ENVRI-Hub represents more than a service and data catalogue; it is a collaborative environment designed to lower barriers for researchers and citizen scientists alike.

Through the ENVRI Catalogue of Services (Figure 1), users can discover and access a diverse range of data access services and analytical tools from atmospheric, marine, terrestrial, and biodiversity domains [1]. This ENVRI-federated approach breaks down traditional silos, a fundamental step for open collaboration, enhancing discoverability and accessibility.

The integrated Virtual Research Environment interface enables researchers to compose, execute, and share analytical workflows using data and services from multiple sources, all within the platform. This not only streamlines complex analyses but also ensures that methodologies and computational environments can be precisely documented and reused, enhancing reproducibility.

Newly developed components, such as the AI Environmental Expert (part of the Knowledge Base) [2], utilise conversational interfaces to assist users in navigating complex scientific resources. This makes the platform's wealth of information more accessible to non-specialists, supporting interdisciplinary inquiry.

A central ENVRI Authentication and Authorisation Infrastructure (AAI) – the ENVRI-ID – enables simplified access to data and metadata sources that require a login, enhancing the user experience with the EOSC-compatible single sign-on system [3].

#### Orientation, Collaboration, and Future Path

The ENVRI-Hub is fundamentally oriented toward community-driven science. Its development has involved close co-creation with a dedicated User Group, ensuring the platform meets real-world needs. Furthermore, its architecture is designed for interoperability with larger European frameworks, such as the European Open Science Cloud (EOSC), positioning it as a key reference in the broader Open Science landscape.

Looking ahead, activities will focus on onboarding more services, expanding user training, and refining the platform based on community feedback. The ultimate goal is to establish the ENVRI-Hub as a sustainable and widely adopted gateway where open, collaborative, and reproducible environmental science becomes the standard practice, empowering researchers to generate the robust evidence needed for science-based policy and a sustainable future.

#### Links:

- [L1] <https://envri-hub.envri.eu/>
- [L2] <https://envri.eu/envri-hub-next/>

#### References:

- [1] ENVRI-Hub NEXT D10.1 Semantic Search in ENVRI Catalogue and RI Catalogues <https://doi.org/10.5281/zenodo.18433332>
- [2] ENVRI-Hub NEXT D9.1 Knowledge Base Recommender System Design <https://doi.org/10.5281/zenodo.18668275>
- [3] ENVRI-Hub NEXT D11.1 Legal Framework for Cross-Domain AAI System <https://doi.org/10.5281/zenodo.18662641>

#### Please contact:

Federico Drago, EGI Foundation, Italy  
federico.drago@egi.eu

Delphine Dobler, Euro-Argo ERIC, France  
delphine.dobler@euro-argo.eu

# D4Science: An Enabling Infrastructure for Open Science

by Massimiliano Assante, Luca Frosini, Francesco Mangiacrapa, and Pasquale Pagano (CNR-ISTI)

**D4Science supports Virtual Research Environments that integrate data, computing, and collaboration tools, making Open Science part of everyday research practice across diverse communities.**

Open Science is increasingly embedded in research practice, but implementing it in everyday workflows remains both a technological and organisational challenge. D4Science [1] addresses this challenge by adopting the as-a-Service paradigm and by offering Virtual Research Environments (VREs) [2], also called Virtual Laboratories (VLabs), as integrated, web-based, working environments. These environments provide researchers with seamless access to data, computational resources, and analytical services within a unified framework. By abstracting away the complexities of storage management, computation, and service orchestration, VREs enable scientists to focus on research design, methodological rigour and knowledge production rather than on IT and infrastructure concerns.

However, fragmentation of tools and the effort required to prepare artefacts for reuse often hinder adoption. D4Science addresses this by embedding Open Science practices directly within VREs.

## An Integrated Research Lifecycle

Each VRE integrates four complementary services that work together as a coherent whole. It includes the Workspace service, a File Sync and Share solution for structured and versioned management of research artefacts, a Collaborative

Research Space that enables contextualised scientific discussion using social networking-like tools, a Data Analytics Platform to execute and share computational workflows where researchers can import and execute their models, and a Catalogue-Based Publishing Platform to disseminate research outputs as early as possible. Together, these components support the entire data-driven research lifecycle, from early hypotheses to analysis and dissemination, within a coherent digital ecosystem.

The Workspace service provides structured, versioned cloud storage for research artefacts. Beyond simple file storage, it assigns persistent identifiers, maintains automatic versioning, and enriches items with extensible metadata. Researchers can keep materials private, share them with collaborators, or make them accessible to the entire VRE community. This flexibility supports progressive openness while preserving control over intellectual contributions.

The Collaborative Research Space enables contextualised interaction within each VRE. Scientific openness also requires transparent communication. Researchers can post, comment on artefacts, annotate results, and discuss methodological choices directly in connection with stored datasets. Discussions are preserved as part of the research record, strengthening transparency and collective validation.

The Data Analytics Platform (see Figure 1) allows researchers to execute computational models in distributed and cloud-based facilities. Algorithms developed in multiple programming languages can be shared and exposed as reusable services. Each analytical execution automatically generates detailed provenance records capturing inputs, parameters, outputs, and execution context. These records are preserved and linked to associated artefacts in the Workspace Cloud Storage. Reproducibility is therefore embedded into routine practice. Publishing a dataset or model also means preserving the computational pathway required to replicate or extend the results.

As scientific knowledge increasingly resides not only in journal articles but also in datasets, software, models, and compu-

Figure 1: The Data Analytics Platform Web User Interface.

tational workflows, the Catalogue-Based Publishing Platform enables researchers to publish these artefacts at different stages of maturity, supporting a genuine “publishing beyond the paper” approach. Each catalogue entry is assigned a persistent identifier and enriched with structured metadata and clear licensing information. This ensures proper attribution, facilitates discovery, and enables reuse. Publication increases visibility within the VRE while also supporting interoperability with external systems through standard interfaces. Publishing is therefore no longer treated as the final step of the research process. Intermediate results, updated versions, and evolving artefacts can be shared early, fostering iterative refinement, transparency, and collaborative improvement.

Recent developments extend these environments with AI-based assistance, as described in the following article by Dell’Amico et al. in this issue.

#### Interoperability and Sustainability

Sustainable Open Science infrastructures require interoperability across platforms and disciplines. D4Science services expose standard interfaces and APIs, enabling integration with external repositories, analytics tools, and research infrastructures. This federated approach reflects the broader EOSC vision of interconnected services operating across institutional and national boundaries. No single platform can serve all scientific domains. Interoperability ensures modular growth and long-term sustainability.

The D4Science experience demonstrates that effective Open Science depends on cohesive ecosystems rather than isolated tools. By reducing fragmentation and lowering the cost of sharing, VREs make openness a natural outcome of daily research activity. Currently, thousands of users [L1] operate across more than two hundred active VREs spanning biodiversity, environmental science, marine research, social sciences, and computational biology. These environments combine domain-specific datasets and services with shared infrastructure components designed to promote FAIR data management and reproducibility. The D4Science model aligns closely with the vision of the European Open Science Cloud (EOSC) [L2], contributing to a federated ecosystem of interoperable services that supports cross-disciplinary collaboration and sustainable digital research infrastructures across Europe.

This alignment is not merely conceptual but already operational. D4Science also forms the technological foundation of the Digital Twin of the Ocean (DTO) EOSC Node [L3], a flagship Thematic Node within the emerging EOSC Federation. Building on the legacy of the Blue-Cloud initiative [3], the DTO Node adopts a 'System of Systems' approach to federate mature marine data infrastructures and modelling services. In this context, D4Science provides the core VRE technology, the federated AAI, and Catalogue interfaces required for seamless integration with the EOSC EU Node. This enables the execution of advanced, cross-domain analytics workflows and the publication of FAIR research artefacts, ensuring that the European marine community is at the forefront of the European digital research landscape.

#### Links:

[L1] <https://www.d4science.org/>

[L2] <https://kwz.me/hIo>

[L3] <https://kwz.me/hIO>

#### References:

- [1] M. Assante et al., “Enacting Open Science by D4Science,” *Future Generation Computer Systems*, vol. 101, pp. 555–563, 2019, doi: 10.1016/j.future.2019.05.063.
- [2] M. Assante et al., “Virtual research environments co-creation: The D4Science experience,” *Concurrency and Computation: Practice and Experience*, vol. 35, no. 18, p. e6925, 2023, doi: 10.1002/cpe.6925.
- [3] D. Schaap, M. Assante, et al., “Blue-Cloud: Exploring and demonstrating the potential of Open Science for ocean sustainability,” in *Proc. 6th Int. Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea)*, IEEE, 2022, pp. 198–202, doi: 10.1109/MetroSea55331.2022.9950819.

#### Please contact:

Massimiliano Assante  
CNR-ISTI, Italy  
[massimiliano.assante@cnr.it](mailto:massimiliano.assante@cnr.it)

## Supporting Open Science in Virtual Research Environments: The DAVE Experience

by Andrea Dell’Amico, Alfredo Oliviero, Giancarlo Panichi, Biagio Peccerillo (CNR-ISTI) and Marco Procaccini (CNR-IGG)

***DAVE, a conversational assistant integrated into D4Science Virtual Research Environments, simplifies access to services and supports Open Science workflows through natural language interaction.***

Virtual Research Environments (VREs) provide integrated access to data, services, and collaboration tools for open and data-intensive research [3].

Building on the D4Science Virtual Research Environments, which support over 230 VREs and around 28,000 users worldwide, our work on conversational agents has led to the development of DAVE (D4Science Assistant for Virtual Research Environments), a system designed to assist researchers directly within their workflows. The D4Science infrastructure and its VREs are described in detail in the article by Assante et al. on page 38 in this issue.

The development of DAVE followed an iterative, user-centered approach that was guided by four key requirements: (a) flexibility and extensibility, as DAVE adopts a highly modular architecture that enables components to be selected, replaced, and extended as needed, ensuring adaptability to an evolving infrastructure; (b) context-awareness, as it relies on a rich and extensible knowledge base that supports heterogeneous research domains and aligns with Open Science ecosystems; (c)

openness and explainability, embodied by the system's transparency in both its behaviour and the knowledge sources behind its responses, which supports trust, reproducibility, and responsible reuse in line with Open Science values; (d) security and trustworthiness, ensuring that privacy and data protection meet the standards of the underlying infrastructure, while also supporting Open Science practices that balance accessibility with the safeguarding of sensitive research data.

Our journey leading to DAVE involved three successive prototypes [1]. The early modular agent, Janet (early 2023), which was based on a pipeline of fine-tuned components, proved to be limited in terms of flexibility and robustness, as well as having a high development and maintenance cost. The second attempt, the D4Science AI Agent (late 2024), adopted the Cheshire Cat framework, moving to a single-agent, multi-tool model. While improving modularity, this architecture proved less suitable for VREs, where diverse services and communities require numerous specialised capabilities. Concentrating all functionalities into one agent led to complex prompt engineering, which increased operational costs and diluted contextual relevance.

DAVE represents a paradigm shift. It adopts a highly modular multi-tool multi-agent system based on the Google Agent Development Kit (ADK) framework. The architecture, shown in Figure 1, features a VRE Assistant acting as a central orchestrator that interprets user requests, plans the sequence of actions, and delegates specific tasks to specialised sub-agents. These specialised agents are tightly integrated with D4Science services within VREs:

1. The Workspace Agent allows users to browse, discover, and summarize scientific documents stored in shared folders;
2. The Catalogue Agent facilitates the discovery and exploitation of research artifacts (datasets, software, publications);
3. The Social Agent summarizes community activities and interactions;
4. The CCP Agent interfaces with the D4Science Cloud Computing Platform, managing the execution of analytics and ensuring the repeatability and reproducibility of research methods.

DAVE simplifies the application of FAIR principles by supporting the discovery and reuse of research artefacts across the scholarly lifecycle. Through natural language interaction, researchers can explore the VRE Catalogue, aggregating datasets, software, and publications, and obtain coherent overviews without navigating complex metadata or multiple interfaces. DAVE also strengthens the collaborative and methodological dimensions of research. By interfacing with the VRE social networking services, it provides contextual

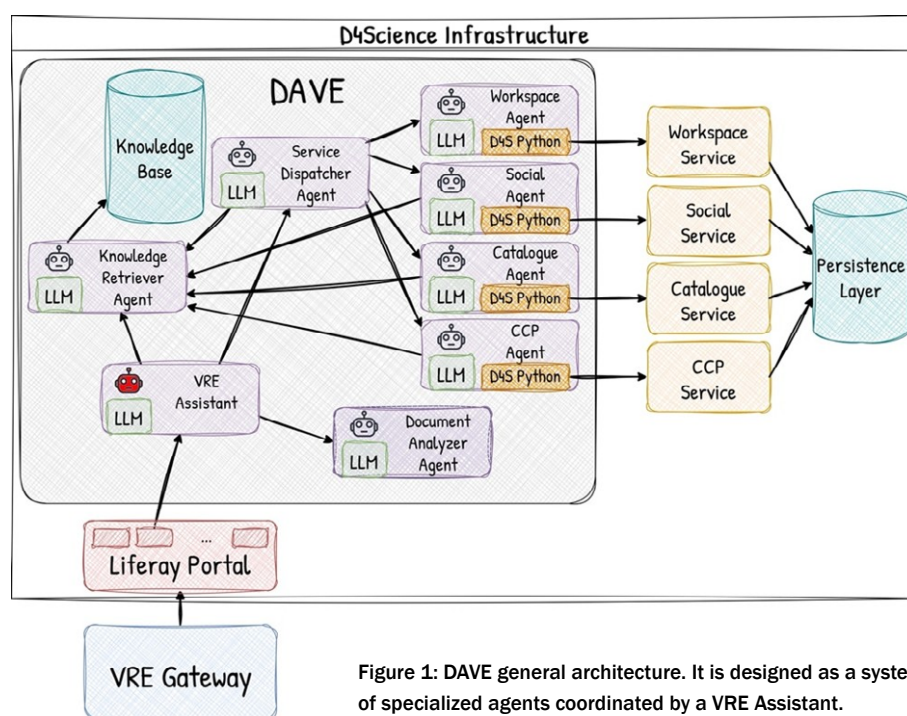


Figure 1: DAVE general architecture. It is designed as a system of specialized agents coordinated by a VRE Assistant.

summaries of community activities and interactions, linking data with ongoing scientific discussion. At the same time, it supports transparency and reproducibility by guiding researchers in preparing, executing, and documenting analytical methods within the Cloud Computing Platform, making algorithms more accessible, executable, and reusable.

In conclusion, DAVE demonstrates how conversational agents can advance Open Science by making VREs more accessible, transparent, and reusable. By supporting the discovery, sharing, and reuse of research artifacts and enabling reproducible methods, DAVE embeds FAIR principles and transparency directly into researchers' workflows, fostering open, trustworthy, and collaborative science.

#### Links:

- [L1] <https://www.d4science.org>  
 [L2] <https://services.d4science.org/web/collab/>

#### References:

- [1] M. Assante et al.: "Deploying Conversational Agents in Virtual Research Environments: Approaches and Lessons Learned", *SN Computer Science*, Volume 7, 2026, art. n. 339.
- [2] M. Assante et al.: "Enacting Open Science by D4Science", *Future Generation Computer Systems*, Volume 101, 2019, Pages 555–563.  
<https://doi.org/10.1016/j.future.2019.05.063>
- [3] L. Candela et al.: "Virtual research environments: an overview and a research agenda", *Data Science Journal*, Volume 12, 2013, Pages GRDI75–GRDI81.  
<https://doi.org/10.2481/dsj.GRDI-013>

#### Please contact:

Biagio Peccerillo, CNR-ISTI, Italy  
[biagio.peccerillo@isti.cnr.it](mailto:biagio.peccerillo@isti.cnr.it)



## HORIZON Europe: ERCIM Project Support Services

A European project can be a richly rewarding tool for pushing your research or innovation activities to the state-of-the-art and beyond. Through ERCIM, our member institutes have participated in more than 100 projects funded by the European Commission in the ICT domain, by carrying out joint research activities while the ERCIM Office successfully manages the complexity of the project administration, finances and outreach.

### HORIZON Europe: How can you get involved?

The ERCIM Office has recognized expertise in a full range of services, including:

- Identification of funding opportunities
- Recruitment of project partners (within ERCIM and through our networks)
- Proposal writing and project negotiation
- Contractual and consortium management
- Communications and systems support
- Organization of attractive events, from team meetings to large scale workshops and conferences
- Support for the dissemination of results.

### How does it work in practice?

Contact the ERCIM Office to present your project idea and a panel of experts within the ERCIM Science Task Group will review your idea and provide recommendations. Based on this feedback, the ERCIM Office will decide whether to commit to help producing your proposal. Note that having at least one ERCIM member involved is mandatory for the ERCIM Office to engage in a project.

If the ERCIM Office expresses its interest to participate, it will assist the project consortium as described above, either as project coordinator or project partner.

### For more information, please contact:

Jessica Michel Assoumou  
ERCIM Office  
[jessica.michel@ercim.eu](mailto:jessica.michel@ercim.eu)



*Europe's Only Summer School  
on Informatics Education Research!*

## Summer School on Informatics Education Research 2026

7-11 September | University of Glasgow, Scotland, UK  
sponsored by IBM Research



For PhD students, post-docs & researchers  
interested in Informatics Education Research

*"It offers a deep dive into the subject with expert instructors, hands-on activities, and the opportunity to network with like-minded professionals, all of which make it an incredibly valuable learning experience."*

*Drishti Yadav, IE Best Dissertation Award 2025 Runner-up  
& Alumni of the 2024 Debut Summer School*

Programme  
Details



**Application Now Opens on a first-come-first-serve basis.**



SCHLOSS DAGSTUHL  
Leibniz-Zentrum für Informatik

Call for Proposals

## Dagstuhl Seminars and Perspectives Workshops

**Schloss Dagstuhl – Leibniz-Zentrum für Informatik is accepting proposals for scientific seminars/workshops in all areas of computer science, in particular also in connection with other fields.**

If accepted, the event will be hosted in the seclusion of Dagstuhl's well known, own, dedicated facilities in Wadern on the western fringe of Germany. Moreover, the Dagstuhl office will assume most of the organisational/ administrative work, and the Dagstuhl scientific staff will support the organizers in preparing, running, and documenting the event. Thanks to subsidies the costs are very low for participants.

Dagstuhl events are typically proposed by a group of three to four outstanding researchers of different affiliations. This organizer team should represent a range of research communities and reflect Dagstuhl's international orientation. More information, in particular details about event form and setup, as well as the proposal form and the proposing process, can be found on

<https://www.dagstuhl.de/dsproposal>

Schloss Dagstuhl – Leibniz-Zentrum für Informatik is funded by the German federal and state government. It pursues a mission of furthering world class research in computer science by facilitating communication and interaction between researchers.

### Important Dates

- *Next submission period:*  
15 October to 1 November 2026
- *Seminar dates:*  
Between October 2027  
and September 2028 (tentative).



## INESC TEC International Visiting Researcher Programme 2026

The INESC TEC has opened applications for the International Visiting Researcher Programme (IIVRP) 2026, offering an excellent opportunity for researchers from around the world to engage in collaborative research within one of Europe's leading applied research institutes.

The programme is designed to attract senior researchers, postdoctoral fellows, and outstanding PhD and Master's students affiliated with institutions outside Portugal. Participants will have the opportunity to join INESC TEC's multidisciplinary research teams and contribute to ongoing projects across a broad range of domains, including artificial intelligence, computer science and engineering, robotics, energy systems, telecommunications, and digital transformation.

Selected candidates will undertake research visits of one to three months, taking place between October 2026 and July 2027. During their stay, visiting researchers will be integrated into active research groups, gaining access to state-of-the-art facilities, international net-

works, and a highly collaborative scientific environment. The programme encourages not only short-term research exchanges but also the development of long-term institutional partnerships and joint research initiatives.

To support participation, the programme offers:

- A monthly allowance of approximately €1,600 to cover living expenses
- A travel allowance, with support varying depending on the country of origin
- Full integration into INESC TEC's research ecosystem, including access to infrastructure and collaboration opportunities.

Beyond individual mobility, the International Visiting Researcher Programme aims to strengthen international cooperation, promote scientific excellence, and facilitate knowledge transfer across borders. It provides a valuable platform for researchers to expand their networks, explore new research directions, and contribute to impactful innovation at the European level.

**Application deadline: 15 May 2026**

Further information and application details: <https://kwz.me/hIY>



## Your Announcement Here

Promote your innovative research or cutting-edge technologies to a global audience of leading researchers, academics, and industry professionals in computer science and applied mathematics through an announcement in ERCIM News.

# EUCNC | 6G Summit

Málaga, Spain, 2-5 June 2026

[www.eucnc.eu](http://www.eucnc.eu)

## ANNOUNCEMENT

### Steering Committee Chairs

Luis M. Correia, IST - U. Lisbon, PT (Chair)  
Pavlos Fournogerakis, SNS JU, BE (Vice-Chair)  
Matti Latva-aho, U. Oulu - 6G Flagship, FI (Vice-Chair)

### Technical Programme Chairs

Pedro Merino-Gómez, U. Malaga, ES  
Ivan Seskar, Rutgers U., US  
Art Pouttu, U. Oulu - 6G Flagship, FI

### Track Co-Chairs

#### PHY - Physical Layer and Fundamentals

Mikko Uusitalo, NOKIA, FI  
Mari Carmen Aguayo-Torres, U. Malaga, ES  
Aggelos Bletsas, Rutgers U., US

#### WOS - Wireless, Optical and Satellite Networks

David Larrabeiti, UCM, ES  
Alessandro Vanelli-Coralli, U. Bologna, IT  
Chia-Pang Yen, ITRI, TW

#### NET - Network Softwareisation

Anna Brunstrom, Karlstad U., SE  
Johann Marquez-Barja, imec-UAntwerp, BE  
Paul Ruth, U. North Carolina, US

#### AIU - Applications, IoT, Use cases

Marc Emmelmann, Fraunhofer FOKUS, DE  
Almudena Díaz, U. Malaga, ES  
Tingjun Chen, Duke U., US

#### OPE - Operational & Experimental Insights

Spyros Denazis, U. Patras, GR  
David Gomez-Barquero, U.P. Valencia, ES  
Aloizio Da Silva, Virginia Tech, US

#### AI4C - AI/ML Solutions for Communications

Daniel Kilper, Trinity College Dublin, IE  
Valerio Frascaola, Intel Deutschland GmbH, DE,  
Jongwon Kim, GIST, KR

#### SAQ - Security Aspects and Quantum Communications

John Preuß Mattsson, Ericsson, SE  
Javier Lopez, U. Malaga, ES  
Vuk Marojevic, Miss. State U., US

#### NVS - Next-Generation Visions & Sustainability

Marja Matinmikko-Blue, U. Oulu - 6G Flagship, FI  
Christoph Schmelz, NOKIA, DE  
Carlos E. Caicedo Bastidas - Syracuse U., US

#### CMA Components, Microelectronics, Photonics & Antennas

Luís Manuel Pessoa, U. Porto - INESC TEC, PT  
Nuutti Tervo, U. Oulu - 6G Flagship, FI  
Ashutosh Dutta, U. Johns Hopkins, US

### Panels Co-Chairs

Håkon Lønsethagen, Telenor, NO  
Özgü Alay, U. Oslo, NO

### Special Sessions Co-Chairs

Čedomir Stefanović, AAU, DK  
Nurul Huda Mahmood, U. Oulu - 6G Flagship, FI  
Filipe Cardoso, Polytech. Inst. Setubal, PT

### Workshops Co-Chairs

Diego López, Telefonica, ES  
Germán Madueño, Keysight Technologies, ES  
Beatriz Lorenzo, U. Mass. Amherst, US

### Tutorials Co-Chairs

Paweł Sroka, PUT, PL  
Mir Ghoraiishi, Gigasys Solutions, UK  
George Sklivanitis, Florida Atlantic U., US

### Exhibitions Co-Chairs

David Artuñedo, Telefonica, ES  
Michael Dieudonne, Keysight Technologies, BE  
Pablo Herrera, U. Malaga, ES

### IEEE ComSoc Liaison

Stefano Bregni, POLIMI, IT

### COST Liaison

Ralph Stübner, COST, BE

### URSI Liaison

Sana Salous, Durham U., UK

### EurAAP Liaison

Jose Molina Garcia Pardo, U. Cartagena, ES

### EURASIP Liaison

Eva Lagunas, U. Luxembourg, LU

The 2026 EuCNC & 6G Summit builds on two successful conferences in the area of telecommunications: EuCNC, in its 35<sup>th</sup> edition of a series, supported by the European Commission, and the 6G Summit, in its 8<sup>th</sup> edition, originated from the 6G Flagship programme in Finland, one of the very first in its area. The conference is sponsored by the IEEE Communications Society (ComSoc), the European Association for Signal Processing (EURASIP) and the European Association on Antennas and Propagation (EurAAP) and supported by the European Commission. The conference addresses various aspects of Beyond 5G/6G communications systems and networks. It brings together cutting-edge research and world-renown industries and businesses, globally attracting in the last years close to 1000 delegates from more than 40 countries all over the world, to present and discuss the latest results, and an exhibition with more than 50 exhibitors, for demonstrating the technology developed in the area, namely within research projects from EU R&I programmes.

The conference program will include:

- Keynotes
- Panels
- Regular Sessions (papers from open call, to be submitted for uploading to IEEE Xplore)
- Special and Convened Sessions, with papers on specific topics
- Workshops, with papers and presentations on specific topics
- Poster sessions
- Tutorials
- Demos and exhibitions, with pitches
- Awards in several categories (paper, poster, booth, demo, startup, thesis)

### Key dates:

- 06 Mar. 2026 – Deadline for submission of Extended Abstracts for Posters
- 09 Mar. 2026 – Notification of acceptance of Workshops, Special Sessions and Tutorials
- 27 Mar. 2026 – Deadline for submission of proposals for Exhibitions, Demos and Booths
- 27 Mar. 2026 – Deadline for submission of applications for Ph.D. Theses and Startups
- 30 Mar. 2026 – Notification of acceptance of Papers and Extended Abstracts
- 10 Apr. 2026 – Deadline for final Papers for all sessions and Workshops
- 13 Apr. 2026 – Notification of acceptance of proposals for Exhibitions, Demos and Booths
- 17 Apr. 2026 – Deadline for Authors registration
- 27 Apr. 2026 – Draft Programme available





ERCIM – the European Research Consortium for Informatics and Mathematics is an organisation dedicated to the advancement of European research and development in information technology and applied mathematics. Its member institutions aim to foster collaborative work within the European research community and to increase co-operation with European industry.



ERCIM is the European Partner of the World Wide Web Consortium.



Consiglio Nazionale delle Ricerche  
Area della Ricerca CNR di Pisa  
Via G. Moruzzi 1, 56124 Pisa, Italy  
[www.iit.cnr.it](http://www.iit.cnr.it)



I.S.I. – Industrial Systems Institute  
Patras Science Park building  
Platani, Patras, Greece, GR-26504  
[www.isi.gr](http://www.isi.gr)



Centrum Wiskunde & Informatica

Centrum Wiskunde & Informatica  
Science Park 123,  
NL-1098 XG Amsterdam, The Netherlands  
[www.cwi.nl](http://www.cwi.nl)



Norwegian University of Science and Technology  
Faculty of Information Technology, Mathematics and Electrical Engineering, N 7491 Trondheim, Norway  
<http://www.ntnu.no/>



Fonds National de la  
Recherche Luxembourg

Fonds National de la Recherche  
6, rue Antoine de Saint-Exupéry, B.P. 1777  
L-1017 Luxembourg-Kirchberg  
[www.fnrlu](http://www.fnrlu)



RISE SICS  
Box 1263,  
SE-164 29 Kista, Sweden  
<http://www.sics.se/>



FORTH  
INSTITUTE OF COMPUTER SCIENCE

Foundation for Research and Technology – Hellas  
Institute of Computer Science  
P.O. Box 1385, GR-71110 Heraklion, Crete, Greece  
[www.ics.forth.gr](http://www.ics.forth.gr)



SBA Research gGmbH  
Floragasse 7, 1040 Wien, Austria  
[www.sba-research.org/](http://www.sba-research.org/)



Fraunhofer ICT Group  
Anna-Louisa-Karsch-Str. 2  
10178 Berlin, Germany  
[www.iuk.fraunhofer.de](http://www.iuk.fraunhofer.de)



Eötvös Loránd Research Network  
Számítástechnikai és Automatizálási Kutató Intézet  
P.O. Box 63, H-1518 Budapest, Hungary  
[www.sztaki.hu/](http://www.sztaki.hu/)



INESC  
c/o INESC Porto, Campus da FEUP,  
Rua Dr. Roberto Frias, n° 378,  
4200-465 Porto, Portugal  
[www.inesc.pt](http://www.inesc.pt)



University of Cyprus  
P.O. Box 20537  
1678 Nicosia, Cyprus  
[www.cs.ucy.ac.cy/](http://www.cs.ucy.ac.cy/)



Institut National de Recherche en Informatique  
et en Automatique  
B.P. 105, F-78153 Le Chesnay, France  
[www.inria.fr](http://www.inria.fr)



UNIVERSIDAD DE MÁLAGA

Institute for Software Engineering and Software Technology  
“Jose María Troya Linero”, University of Malaga  
Calle Arquitecto Francisco Peñalosa, 18, 29010 Málaga  
<https://gp.uma.es/itis>