

# Together Building Open Access



CO<sup>n</sup>necting RE<sup>po</sup>sitories



# Unlocking Global Knowledge

## NEWSLETTER

March 2026



<https://core.ac.uk/governance>

# Content

<b>1.Introductory Letter</b> .....	4
<b>2.Infrastructure &amp; Technology Developments</b> .....	5
2.1CORE Infrastructure Migration .....	5
<b>3.Repository Tools &amp; Compliance</b> .....	5
3.1Integrating CORE Functionality with DSpace and DSpace-CRIS .....	5
3.2FAIR Certification Module .....	6
3.3Updating the CORE Compliance Tool for REF 2029 .....	7
<b>4.Community, Collaboration &amp; Networks</b> .....	8
4.1Educational Series: Making the Most of the CORE Dashboard .....	8
4.2Horizon Europe grant .....	9
<b>5.Partnerships &amp; Sustainability</b> .....	11
5.1SoFAIR Project: Final Outcomes and Reflections .....	11
<b>6.Conferences, Advocacy &amp; External Engagement</b> .....	12
6.1Open Repositories 2026: CORE Submission .....	12
<b>7. Additional Reading – Opinion Piece: The Case for Authoritative Custodianship of Scholarly Metadata</b> .....	13

# Welcome to Our New CORE Members 🎉



BUCKINGHAMSHIRE  
NEW UNIVERSITY  
EST. 1891

# Thank You for Renewing Your Membership



University  
of Exeter



Brunel  
University  
London



Nottingham Trent  
University



UNIVERSITY OF  
BIRMINGHAM



University of  
BRISTOL



Imperial College  
London

US

UNIVERSITY  
OF SUSSEX



White Rose  
Libraries  
Universities of Leeds, Sheffield & York



Lancaster  
University

# Our Thanks to Every CORE Member



UNIVERSITY OF  
CAMBRIDGE

ual:



UNIVERSITY OF  
OXFORD



University of  
Nottingham



Manchester  
Metropolitan  
University



UNIVERSITY OF  
BRADFORD



TIB  
LEIBNIZ INFORMATION CENTRE  
FOR SCIENCE AND TECHNOLOGY  
UNIVERSITY LIBRARY



University of  
Reading



Queen Mary  
University of London



CARDIFF  
UNIVERSITY  
PRIFYSGOL  
CAERDYDD



TU  
WIEN



University of  
Strathclyde



St George's  
University of London

CITY  
UNIVERSITY OF LONDON  
EST 1894

CITY  
UNIVERSITY OF LONDON  
EST 1894



THE UNIVERSITY OF  
CHICAGO



UNIVERSITY  
OF HULL



UNIVERSITY OF  
GREENWICH



University  
of Glasgow



Durham  
University



UNIVERSITY OF  
BATH

# Dear Members of the CORE Community,

Scholarly communication continues to undergo significant change, shaped by automation, artificial intelligence, and an increasing importance of shared infrastructures. These shifts are reshaping how research outputs are produced, described, and reused. In this context, the question of stewardship, who holds responsibility for the scholarly record has never been more pressing.

Academic institutions are, at their core, knowledge-producing organisations. The research outputs created by their staff represent their most valuable and enduring assets. Yet many institutions today struggle to maintain a complete, authoritative account of what they have produced, increasingly depending on third-party systems to reconstruct their own scholarly record.

My position on this, which has long guided CORE's work, is clear: institutions must be able to retain authoritative custodianship of the metadata describing their research outputs. Authoritative custodianship is a strategic asset that makes the open repositories network so valuable and resilient. Institutional repositories are the natural systems of record for this role. While institutional resources might be under constraint, we need to find a way, as a community, to enable repositories to preserve this role. For readers interested in exploring this position in more detail, a longer opinion piece expanding on these ideas is included at the end of this newsletter.

We are also pleased to share that CORE has two accepted presentations for **Open Repositories 2026**, taking place fully online from 8–11 June 2026 and we an accepted submission for **FORCE 2026** in Singapore in from 2-5 June 2026:

- *Towards the integration of CORE into DSpace* – exploring collaboration with 4Science to embed CORE functionality in repository platforms, enhancing interoperability and research visibility.
- *Large Language Models for Software Mention Extraction* – highlighting innovative AI approaches to automatically identify and link research software to the papers that introduced it, supporting reproducibility and research transparency.
- *Managing AI Bot Access to Open Scholarly Infrastructures* - this reports on the issues and recent work of the CORE team in the area of protecting scholarly infrastructures from abusive AI bots while retaining access for mission-aligned services.

This newsletter features latest improvements within the CORE platform in a range of areas, including: technical integrations with repository platforms, recent improvements in compliance and FAIR assessment tools, new guidelines for OA compliance in preparation for REF 2029, our latest recommendations on metadata governance in institutional repositories, and how community collaboration, through initiatives such as training and educational resources, continues to strengthen repositories and support the teams that manage them.

Over the past year, we have also taken time to clearly articulate [CORE's Distinctive Characteristics](#), defining what is unique about CORE, how it influences our positioning among global open scholarly infrastructures. We believe that this is an important step in being even more transparent about how we work, what we prioritise, and the values that shape our infrastructure. Much of this has long guided our decisions and collaborations; making it explicit helps our community better understand how CORE is built, governed, and sustained. It also reflects how community collaboration continues to shape our direction, reminding us that sustainable infrastructure is built not only through technology, but through shared values and collective responsibility.

CORE's future, like its past, depends on partnership with the community it serves. Stewardship cannot be outsourced; it must be practised, supported, and continuously reaffirmed. We remain committed to working with you to ensure that institutional repositories continue to serve as trusted systems of record, enabling open discovery while safeguarding the integrity of the scholarly record for generations to come.

With thanks for your continued trust and collaboration.



**Petr Knoth**

**Founder and Head of CORE**

## 2. Infrastructure & Technology Developments

### 2.1 CORE Infrastructure Migration

As we announced in the October 2025 Newsletter, CORE is now in the final stages of a migration process that will see all of CORE's infrastructure resources moved to the cloud. This is to ensure scalability and to increase the robustness, reliability and resilience of the system. At this stage, about 75% of CORE functionalities have already been migrated and only our indexing continues to run from our on-premise hardware. We anticipate completion in the next few months in accordance with our original plans. We hope this will elevate CORE's infrastructure to a new level and make it easier to scale our services, support new use cases, and provide more reliable access to the global collection of open research.



## 3. Repository Tools & Compliance

### 3.1 Integrating CORE Functionality with DSpace and DSpace-CRIS: Project Updates

We recently met with the 4Science DSpace team to chart the initial steps for integrating key CORE services. This is a phased approach focused on bringing new utility to our shared users.

#### Current Focus: Initial Integration Priorities

Here is what we've prioritized for the first phase:

#### CORE Search API Integration

This will enable users to import publication metadata directly from CORE into new DSpace submissions.

#### OAI identifier/DOI integration

Adding OAI identifier as an entity in the "Identifiers" field during submission process. That will allow users to prefill metadata by inserting OAI.

#### Enrichment: SDG integration

Import information about SDGs identified by CORE in batch: to automatically add SDG metadata to the item as identified by CORE or suggest (mediated by repository admins) the addition of such metadata.

#### CORE Recommender

Recommender to be integrated in the generic DSpace item page and/or for specific entities (i.e. publication), the list of records recommended by CORE related to the visited item.

By leveraging our respective expertise, we envision a strategic partnership that strengthens both our service offerings and the overall research infrastructure. This collaboration represents an important step toward making repositories smarter, more connected, and more impactful.

## 3.2 FAIR Certification Module

CORE has developed a **FAIR Certification Module**, designed to help repositories evidence alignment with the FAIR principles through a transparent, data-driven, and technology-supported process. The module assesses repositories against objective criteria covering findability, accessibility, interoperability, and reusability, providing clear feedback and practical recommendations for improvement.

A prototype of the module has now been implemented and will be showcased to a selected group of repositories at the end of March, where feedback will be collected to inform further refinement prior to wider release. The anticipated public release is summer 2026.

FAIR Certification will offer repositories a structured way to demonstrate good stewardship, improve machine-actionability, and align with best practice in open research infrastructure. Certification will be available at **Bronze, Silver, and Gold levels**, reflecting increasing FAIR maturity and technical capability, and will be supported directly through the **CORE Dashboard**. For **Sustaining Members**, certification will be available at no additional cost.

The module builds on CORE's long-standing role in supporting repository quality, metadata governance, and open, non-discriminatory indexing, helping repositories remain authoritative systems of record while maximising the visibility and reuse of the research they host.



### Bronze

Guarantees a foundational level of compliance. Repository metadata and content is indexable, records are assigned identifiers and metadata have been populated to at least a minimum level.



### Silver

Demonstrates strong compliance with FAIR. Metadata are exposed using industry-leading application profiles and vocabularies, contain machine-readable licence information and have been sufficiently populated.



### Gold

Achieves full machine-actionability. Metadata have been richly populated and support linking of research papers and datasets. Repository supports industry-leading protocols and interoperability standards.

[About CORE FAIR Certification](#)

[Download report](#)

[Download badge in PNG](#)

[Download certificate in PDF](#)

## FAIR certification: Bronze

*In the report below you can review the estimation to what extent your repository correspond to the FAIR principles.*

**Certificate issued:** 26.05.2025

**Certificate valid until:** 26.05.2025

**Date of the last report update:** 26.05.2025



### 3.3 Updating the CORE Compliance Tool for REF 2029

# REF 2029

Research Excellence Framework

As UK institutions begin preparing for REF 2029, CORE continues to support members in navigating the updated open access policy requirements and audit expectations. We have updated the REF 2029 guidance on [this page](#), including changes to deposit timing, embargo periods, licensing expectations, and audit interpretation. These updates are designed to help repository teams monitor compliance early, identify risks, and improve metadata quality well ahead of submission, reducing uncertainty and last-minute remediation.

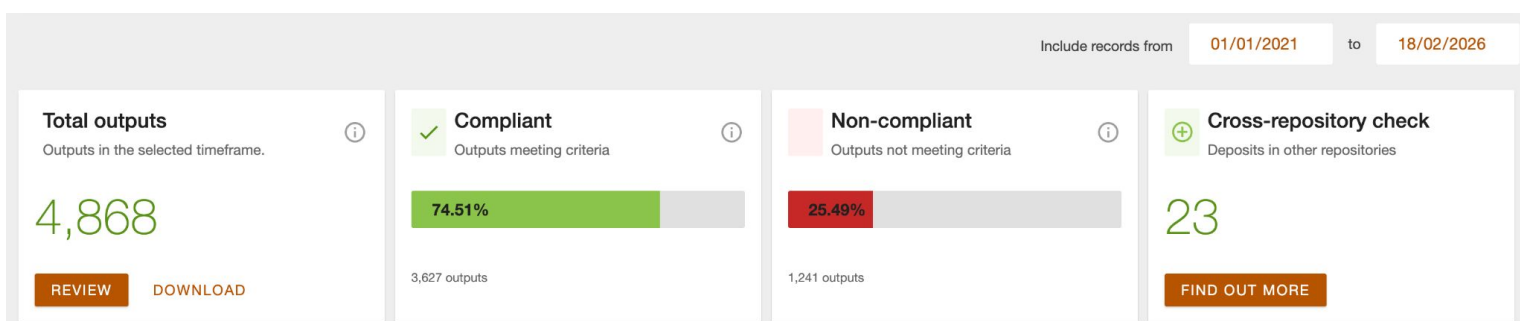
CORE's role is not to replace institutional responsibility, but to support it providing sector-wide visibility, consistent interpretation of policy requirements, and practical feedback through the CORE Dashboard. Alongside the tool updates, we are engaging directly with the REF team and will continue to share guidance as audit processes are clarified.

In addition, we are **organising a single-session working group at the end of March** to discuss these updates and their practical implications. This will be an opportunity for repository teams to ask questions, share experiences, and discuss strategies for compliance and metadata management.

In [CORE Recommendations](#), we cover the following topics:

- Significant changes in REF2029 Open Access Policy
- REF2029 Audit – Guidance
- Recommendations for exposing research outputs' metadata to CORE for REF2029
- How CORE indexing supports HEIs
- How CORE collects this information
- How CORE can help you

We offer a **free REF review and analysis for CORE Sustaining Members**. Additionally, the **CORE Dashboard**, available for all data providers, provides useful information to check that your repository is being properly indexed.



Is your institution correctly set up for REF2029?

[Contact us](#)

## 4. Community, Collaboration & Networks

### *Reinforcing CORE as a community-powered infrastructure*

#### 4.1 Educational Series: Making the Most of the CORE Dashboard

Following the recent milestone of 15 years of supporting open scholarly infrastructure, CORE has begun an educational blog and email series designed to help repository managers better understand and make full use of the **CORE Dashboard**.

Each instalment explores a specific area of the dashboard, explaining what the data represents and how it can support repository visibility, metadata quality, and open access compliance workflows. The series is intended as a practical guide for repository teams who want to better understand how their content is represented in CORE and how the available tools can support day-to-day repository management.

The series began with the **Overview tab** – the first page visible when logging into the dashboard and the quickest way to understand how your repository is currently represented within CORE. You can read the first instalment here:

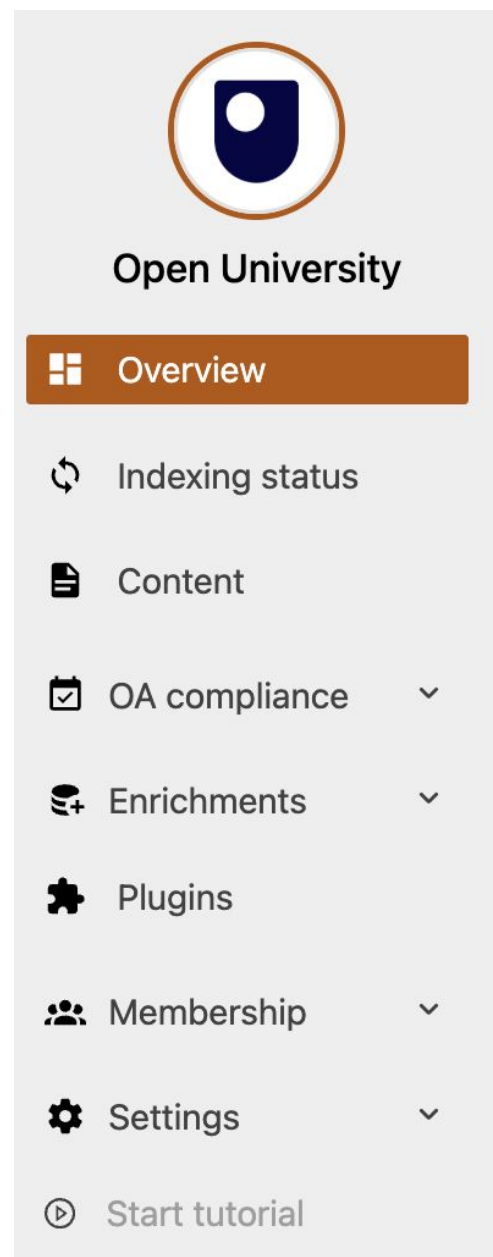
[Understanding the CORE Dashboard: A Closer Look at the Overview Tab](#)

Over the coming months, the series will continue with short, focused guides covering additional areas of the dashboard, including:

- Overview
- Indexing Status
- Content
- OA Compliance: REF 2029
- OA Compliance: Metadata Validator
- OA Compliance: Desirable Characteristics Report
- Enrichments: DOI, ORCID, SDG Insights, RRS Policy, DAS Identification, Research Software, and Versions/Duplicates
- Plugins
- Membership
- Settings: Organisational, Repository, and Notifications

To accompany the series, we will also host informal 30-minute open Teams sessions where repository managers can join, ask questions, and discuss the topics covered in the guides with the CORE team. These sessions are optional and intended as a supportive space for anyone who would like clarification or practical advice while exploring the dashboard.

This initiative reflects CORE's ongoing commitment to working closely with our community and ensuring that repository teams have the knowledge and tools needed to maximise the visibility, quality, and impact of the research they steward.



# Professor Petr Knoth secures €1.37m Horizon Europe grant as part of €9m consortium developing AI research assistants for the European Open Science Cloud

Professor Petr Knoth, Head of CORE at The Open University, has secured **€1.37 million in Horizon Europe funding** as part of a **€9 million international consortium** that will develop new artificial intelligence assistants to support research within the **European Open Science Cloud (EOSC)**.

The project, **EOSC AIAssistant**, is coordinated by **TIB – Leibniz Information Centre for Science and Technology**, who are also CORE Sustaining Members. The project brings together leading European research infrastructures and organisations to explore how emerging *agentic AI*

systems can support researchers in navigating literature, connecting datasets and carrying out complex multi-step research tasks.

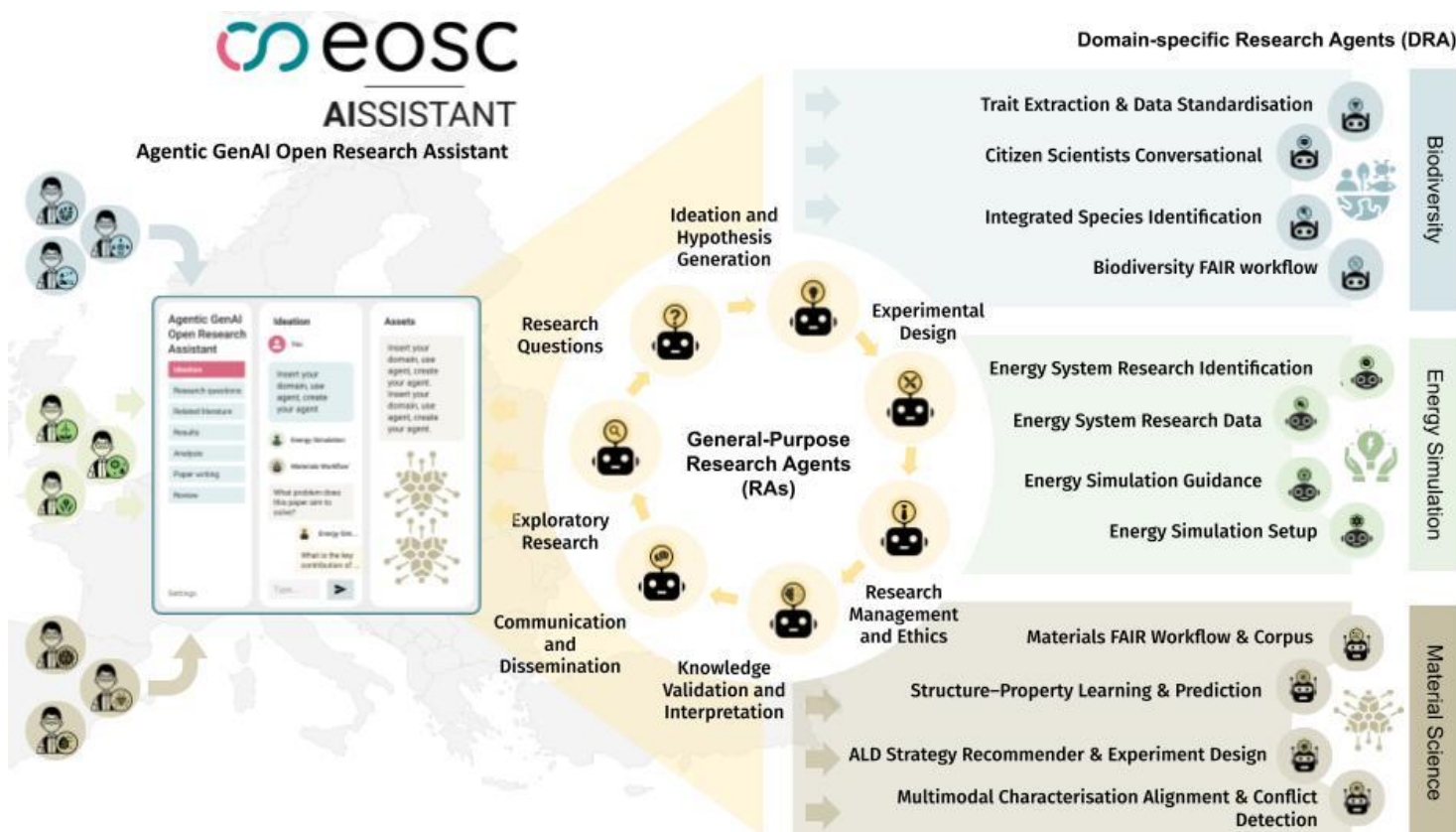
Professor Knoth and his team represent the **largest partner in the consortium after the coordinator** and will lead research worth €1.37 million over the next three years.

A central objective of the project is to ensure that AI systems used in research are **ethical, responsible and aligned with the norms of scholarly communication**. EOSC AIAssistant places strong emphasis on transparency, traceability and proper referencing of scientific sources so that AI-generated outputs remain verifiable and grounded in trusted research.

## AI agents for research workflows

The project will develop a range of **AI research agents** designed to support scientists in their day-to-day research activities.

Some of these will be **domain-agnostic agents**, capable of supporting common research tasks across disciplines. Examples include agents that assist with literature discovery and review, help researchers explore connections between publications and datasets, and support structured evidence gathering across large bodies of scientific literature.



Alongside these, the project will also develop domain-specific agents tailored to particular scientific fields. These will be tested with expert partners working in areas such as biodiversity, energy systems and materials science to understand how AI-assisted research workflows operate in different research environments.

Professor Knoth and his team will contribute both as a **key provider of scholarly data** and by **leading the development of several of the domain-agnostic AI research agents**.

### The role of CORE

CORE will play an important role in the project by providing large-scale scholarly data and expertise in managing and analysing research information.

As **the largest open access research aggregation services**, CORE indexes millions of research papers from repositories and journals worldwide, providing machine-readable metadata and full text suitable for large-scale text and data mining.

Within EOSC Allisstant, these datasets will help ensure that AI systems are grounded in **verifiable scientific knowledge** and can properly **reference and acknowledge the original research sources** they rely on.

For repositories contributing content to CORE, this also means that their research outputs are well positioned to become discoverable within the next generation of **AI-powered research tools being developed for the European Open Science Cloud**.

### Looking ahead

EOSC Allisstant will run for three years and contribute to the broader vision of EOSC as a federated environment where researchers can access data, publications, services and computational tools to support open and collaborative science.

Professor Petr Knoth commented:

*“We are excited to contribute to the development of AI systems that support researchers while remaining grounded in trusted scholarly sources. Ensuring that AI tools properly reference and acknowledge the research they rely on is essential for maintaining the integrity of the scholarly ecosystem.”*



## 5. Partnerships & Sustainability

### 5.1 SOFAIR Project: Final Outcomes and Reflections



The final SoFAIR webinar, held online on 11 December 2025, presented the goals, methodology, and outcomes of the SoFAIR project (Making Software FAIR: A machine-assisted workflow for the research software lifecycle). SoFAIR is a two-year CHIST-ERA funded project, led by The Open University, and involving six partners across four countries. The SoFAIR project partners The project's overarching goal is addressing persistent challenges in research software discoverability, attribution, and long-term preservation. Research software is frequently mentioned only implicitly in scholarly articles, preventing it from becoming a first-class, FAIR (Findable, Accessible, Interoperable, Reusable) research object.

The outcomes from the SoFAIR project include a scalable, machine-assisted workflow that automatically identifies software mentions in full-text articles, and then provides a mechanism for author validation of these mentions. CORE, as a nexus for the global repositories network, is ideally placed to extract these mentions and then route the author validation requests to the relevant repository via a module in the CORE dashboard. Further, once validated, the software metadata is linked to the paper's metadata and the software is archived by Software Heritage (one of the project partners), and a permanent identifier is generated. SoFAIR builds on existing standards and repository practices, notably the COAR Notify protocol which is the communication link between each of the SoFAIR modules.

This ensures the outputs of the SoFAIR project are extensible and can be integrated into repository systems in a number of different ways. Some of these were covered during the webinar which showcased three use cases that have been completed during the lifespan of the project; EuropePMC is integrating software mentions generated as part of the SoFAIR project and including these mentions in the metadata of published articles.

The HAL Open Science repository developed a 'short loop' for authors within the HAL ecosystem, allowing for the validation of software mentions within the HAL repository platform itself. Finally CORE worked closely with Software Heritage to make sure that software mentions are shared widely in the research graph, promoting deposit and help tracking connections between research software and the papers which first introduced it. The project workflow has been publicly released in the form of documentation, software and data via the SoFAIR project's website and workflow and solution documentation pages and the project's Github repository.

The webinar also included a panel discussion and we're extremely grateful to our panel members; Arfon Smith – Schmitt Sciences, Kathleen Sheerer – Executive Director of COAR (Confederation of Open Access Repositories) and John Salter, Research Library Technician at White Rose Libraries, University of Leeds. The discussion brought together perspectives spanning research infrastructure coordination, repository operations, and policy-level stewardship. Chaired by Professor Petr Knoth (CORE, The Open University), the discussion was framed around the practical and cultural conditions required to make research software genuinely reusable and trustworthy. You can watch a recording of the webinar on the CORE Youtube channel.

[Read more about SoFAIR final webinar](#)

[Blog post](#)

# OPEN REPOSITORIES 2026

Online conference | 8-11 June 2026

### 6.1 OR2026 Presentation Submission

21st International Open Repositories Conference, 8-11 June 2026 Fully Online

#### Towards the integration of CORE into DSpace

Viktoriia Pavlenko, CORE,  
[viktoriia.pavlenko@open.ac.uk](mailto:viktoriia.pavlenko@open.ac.uk)

Matteo Cancellieri, CORE, KMi, The Open University.  
[matteo.cancellieri@open.ac.uk](mailto:matteo.cancellieri@open.ac.uk)

Giuseppe Digilio, 4Science,  
[giuseppe.digilio@4science.com](mailto:giuseppe.digilio@4science.com)

Andrea Bollini, 4Science,  
[andrea.bollini@4Science.com](mailto:andrea.bollini@4Science.com)

Petr Knoth, CORE, KMi, The Open University.  
[petr.knoth@open.ac.uk](mailto:petr.knoth@open.ac.uk)

#### Abstract

This proposal outlines key areas for collaboration where 4Science's advanced repository solutions, CRIS systems, and data management capabilities can integrate with CORE's vast aggregation of open access content to maximise research visibility, interoperability, and impact.

#### Large Language Models for Software Mention Extraction

David Pride, CORE, Knowledge Media institute, The Open University,  
[david.pride@open.ac.uk](mailto:david.pride@open.ac.uk)

Matteo Guenci, University of Bologna,  
[matteo.guenci@studio.unibo.it](mailto:matteo.guenci@studio.unibo.it)

Martin Dočekal, Brno University of Technology, [idocekal@fit.vutbr.cz](mailto:idocekal@fit.vutbr.cz)

Silvio Peroni, University of Bologna, [silvio.peroni@unibo.it](mailto:silvio.peroni@unibo.it)

Petr Knoth, CORE, Knowledge Media institute, The Open University,  
[petr.knoth@open.ac.uk](mailto:petr.knoth@open.ac.uk)

#### Abstract

A large proportion of scientific studies now rely on software and data as an integral component of the research process. Significant time and resources are committed to the development of research software yet, too often, these valuable assets lie languishing, hidden in the original research paper that presented them. Ensuring the availability of software and data, and directly linking these assets to the research that first introduced them, is a key component in addressing current problems faced by many scientists when attempting to replicate earlier studies. There have been a number of efforts in recent years to develop methodologies for the extraction and classification of software mentions found in full text scholarly documents. In this presentation, we will discuss how large language models can match current SotA approaches to the problem utilising zero-shot methods that require no pre-training.

## 7. Additional Reading - Opinion Piece

We would like to hear your feedback about the below position paper draft. Do you agree, do you disagree and if so why? If you have an opinion, please reach out directly to [petr.knoth@open.ac.uk](mailto:petr.knoth@open.ac.uk)

### **Authoritative Custodianship of Scholarly Metadata in the Age of AI: Why Institutions Must Retain Control While Enabling Open and FAIR Indexing**

Petr Knoth, CORE, KMi, The Open University, [petr.knoth@open.ac.uk](mailto:petr.knoth@open.ac.uk) ORCID: [0000-0003-1161-7359](https://orcid.org/0000-0003-1161-7359)

#### **Abstract**

Academic institutions are knowledge-producing organisations whose most valuable and enduring outputs are research publications authored by their staff. Yet many institutions are unable to provide a comprehensive and authoritative account of their scholarly production without relying on external systems. We argue that maintaining an authoritative, comprehensive, and institutionally governed record of research outputs is a core institutional responsibility that should not be externalised. We introduce the concept of *authoritative custodianship of metadata* and argue that institutional repositories are the natural systems of record for fulfilling this role. This position is explicitly supportive of a pluralistic ecosystem of indexing systems, which depend on open and non-discriminatory access to repository metadata. We distinguish repository indexing (enabled e.g. by OAI-PMH) from metadata delegation to identifier registries and question the assumption that long-term custodianship can be permanently outsourced through pay-for-registration models. In the age of artificial intelligence, where research outputs increasingly function as high-value and trustworthy knowledge sources, institutional stewardship of authoritative metadata becomes strategically essential for provenance, accountability, and fair access to scholarly knowledge. This paper is presented as a position paper for the Open Repositories community, aiming to stimulate reflection and discussion on the roles of institutions, repositories, indexing systems, and metadata registration agencies in a rapidly changing technological landscape.

#### **Keywords**

*Custodianship, Authoritative metadata management, AI, preservation, indexing, discoverability, institutional responsibility*

#### **Audience**

*The primary audience for this presentation is repository managers, librarians, institutional research leaders, and scholarly infrastructure practitioners within the Open Repositories community who shape how institutional research outputs are curated, exposed, and reused at scale.*

#### **Time Zone**

GMT+1

## **1. Introduction**

Universities and research institutions are, at their core, knowledge organisations. Their primary purpose is the creation, validation, and dissemination of knowledge through research and scholarship. Scholarly publications represent the most visible, durable, and valuable outputs of this activity, embodying intellectual labour, institutional investment, and public funding.

Despite this, many academic institutions are unable to answer a seemingly basic question with confidence: what research outputs has this institution produced? Instead, institutions increasingly rely on third-party systems, such as bibliographic databases, analytics platforms, or identifier registries, to reconstruct their own scholarly record.

This paper argues that this situation reflects a gradual erosion of institutional responsibility rather than a technical necessity. We contend that academic institutions should retain authoritative custodianship of metadata describing their research outputs. Institutional repositories already provide a safe, sustainable, and appropriate infrastructure for fulfilling this role. While indexing systems are essential for discovery and reuse, they are inherently consumers of institutional metadata, and their role is not to replace institutions as the primary sources of truth.

## **2. Research Outputs as Key Institutional Assets**

Research outputs are not incidental by-products of academic activity. They are among the most valuable assets universities create, designed to persist for centuries and to contribute to the cumulative scholarly record. Publications underpin institutional reputation, research assessment, funding allocation, and public trust in science.

From an organisational perspective, it follows that institutions should maintain a complete and authoritative understanding of what they have produced. In most sectors, organisations would not consider it acceptable to lack an internal system of record for their most valuable intellectual assets. Yet within academia, responsibility for tracking and describing research outputs is often perceived as something that can be safely externalised.

This perception has been reinforced by the establishment of identifier registration entities initially created by publishers to manage scholarly information. As a result, many repositories mistakenly believe that their primary focus should be on ensuring all metadata assets within a repository are registered with these agencies. A contributing factor to this belief is the misconception among some academics that a piece of research is not considered valuable unless it has a DOI. This perspective overlooks the fundamental purpose of repositories, which is to authoritatively expose metadata and facilitate access and sharing within a distributed and resilient network of open repositories. Rather than demonstrating success, this situation highlights a historical burden on the scholarly community, emphasising legacy concerns over the actual functionality of repositories.

## **3. The Normalisation of External Dependency**

Many institutions have, over time, become dependent on external systems to manage and make sense of their own scholarly assets, and are increasingly recognising that they lack direct awareness of all the research outputs created by their academics. This dependency can hinder their ability to assess the full scope of their research contributions, exercise institutional oversight, and adequately showcase the work of their university. While external systems may offer valuable services, reliance on them as de facto systems of record highlights the need for institutions to re-establish internal mechanisms that allow them to track, understand, and authoritatively curate their own research outputs.

This dependency has several consequences. First, institutions lose authoritative control over the metadata describing their scholarly outputs and how this metadata is represented and maintained, which in turn means that institutional leadership no longer has a complete and authoritative view of the most valuable knowledge assets the organisation creates. Second, institutions become dependent on infrastructures whose governance models, funding assumptions, and sustainability horizons may not align with those of academic institutions, or whose operational models may change in the future. Third, the institution's own repository risks being perceived as incomplete or secondary, further reinforcing a cycle of external dependency, weakening internal knowledge governance and the case for strategic institutional support.

This situation is not uniform across regions. In the UK, for example, policy drivers such as the Research Excellence Framework (REF) have led to high levels of repository deposit and near-comprehensive institutional records. In other contexts, including parts of the United States, repositories are often less complete, further entrenching reliance on third-party systems. These differences suggest that repository comprehensiveness is achievable where institutions choose to prioritise it.

#### **4. Authoritative Custodianship of Metadata**

This position aligns closely with recent community-led initiatives advocating for openness and transparency in scholarly metadata, most notably the Barcelona Declaration on Open Research Information [1]. The Declaration calls for research information, including metadata about scholarly outputs, to be openly available, interoperable, and governed in the public interest. Importantly, it emphasises that openness of metadata should not imply a loss of institutional responsibility or control, but rather requires that authoritative sources of research information remain visible, accessible, and reusable.

At the heart of this paper is a distinction between access to metadata and custodianship of metadata.

We define *authoritative custodianship of metadata* as responsibility for maintaining the primary, complete, and governance-backed record of metadata describing an institution's research outputs. Custodianship entails accountability for accuracy, continuity, provenance, and long-term stewardship.

Indexing and aggregation systems require broad access to metadata in order to index, analyse and facilitate the discoverability of research outputs and knowledge. Custodianship, by contrast, concerns who is responsible for maintaining the authoritative version of that metadata. We argue that while metadata should be openly accessible to any well-behaved agent, authoritative custodianship must remain with the institution itself and should not be delegated or externalised.

#### **5. Institutional Repositories as Systems of Record**

Institutional repositories were established to provide stable, institutionally governed infrastructures for collecting, curating, preserving, and exposing scholarly outputs. They are operated under institutional control, aligned with institutional longevity, and embedded within local governance and preservation frameworks. From this perspective, repositories should be understood as systems of record rather than merely dissemination platforms. They provide a natural locus for asserting provenance, maintaining authoritative metadata, and enabling both human and machine access to institutional knowledge.

That repositories are sometimes incomplete should be understood as a matter of incentives and institutional practice, not as evidence of their unsuitability. When repositories are treated as authoritative systems of record, they support both institutional accountability and downstream reuse.

#### **6. Open and Non-discriminatory Indexing: Access Without Privilege**

Large-scale indexing and processing of research outputs is essential for discoverability, accessibility and innovation streamlining. No single repository can, itself, provide the cross-institutional view required by contemporary research communities.

Institutional repositories should therefore be openly and non-discriminatorily accessible to indexing systems (including AI bots), in the same way that the open web is accessible to search engines. Provided machine agents behave responsibly, there is no principled reason to restrict their ability to visit open repositories and index research outputs. Broad and fair indexability is a prerequisite for an open and competitive ecosystem.

Crucially, indexing systems read institutional metadata; they do not assume custodianship over it. A healthy infrastructure is one in which indexing systems visit repositories under transparent conditions, rather than one in which bibliographic creation is granted as a privilege to a limited set of registration actors.

## 7. Identifier Registries and the Limits of Metadata Delegation

It is important to distinguish indexing systems from identifier registries such as Crossref and DataCite. These registries play a valuable role in identification and linking and they currently operate in good faith, following recognised governance principles, including POSI and living will commitments.

The issue discussed in this paper is not related to the behaviour of these organisations, but rather to the structural assumptions that form the basis of their business model. The mentioned Identifier registries (as opposed to free of charge and decentralised persistent identification systems, such as Ark[2] and OAI[3]) require institutions to transfer and register metadata under a pay-per-identifier model, which implicitly assumes persistent stewardship and resolution over very long time horizons.

From an institutional perspective, the assumption that metadata custodianship can be permanently outsourced through such models is difficult to reconcile with the centuries-long timescales over which universities operate. Many academic institutions have existed for centuries, whereas identifier registries are comparatively recent entities. This raises the question of whether long-term persistence can realistically be guaranteed, even with the best of intentions, through this model, and whether custodianship of scholarly assets should ultimately rest with the institutions that created them or with external services.

**A pay-at-the-time-of-registration model relies on the continued influx of new registrations to sustain the infrastructure** required for persistent resolution of identifiers and stewardship. As a result, its promise of persistence becomes fragile if participation declines or priorities change. While such models may function effectively for extended periods, they remain contingent on perpetual financial contributions flowing towards a specific third-party provider, resulting in a vendor lock-in for institutions. The model is based on unrealistic conditions that are increasingly difficult to guarantee in a world marked by geopolitical instability, economic shocks, and rapid technological change. Identifier registries should therefore be understood as service providers that consume and resolve institutional metadata at present and in the mid-term future, but not as substitutes for institutional systems of record.

## 8. Institutional Stewardship in the Age of AI

The emergence of large-scale artificial intelligence systems, including Large Language Models (LLMs), introduces both significant risks and important opportunities for scholarly communication. On the one hand, LLMs amplify the risk of misinformation and decontextualised knowledge when trained on opaque or low-quality sources. On the other hand, they create unprecedented opportunities to improve access to research findings, support evidence-based question-answering, and make scientific knowledge more accessible than ever to the general public.

The key to navigating this tension is **provenance**. Research outputs, when accompanied by authoritative metadata that clearly establishes origin, authorship and institutional context, provide uniquely trustworthy inputs for AI systems. Ensuring such provenance cannot be delegated entirely to external parties.

It requires that institutions take responsibility for the knowledge assets their academics create and act as authoritative custodians of the metadata and content that underpin responsible AI use.

Compared to much of the open web, scholarly publications offer structured authorship, institutional context, and attribution. These characteristics make them uniquely valuable as building blocks for AI systems. Preserving these qualities requires that institutions retain authoritative control over how their scholarly outputs are curated and described.

In this context, institutional repositories become strategic assets. By exposing authoritative, machine-readable metadata under open conditions, repositories enable a plural ecosystem of indexing and AI-driven services to emerge, rather than concentrating epistemic power within a small number of intermediaries.

## 9. Implications and Conclusion

Academic institutions produce knowledge, and research outputs are among their most valuable and enduring assets. Maintaining an authoritative, comprehensive record of these outputs is therefore a core institutional responsibility. This paper argues for a clear and fair division of responsibility: institutions as authoritative custodians of metadata, repositories as systems of record, and indexing systems as open, non-discriminatory consumers of that metadata. Identifier registries provide valuable services, but should not be treated as replacements for institutional stewardship.

In the age of AI, where research outputs function as high-value and trustworthy inputs to automated systems, retaining authoritative custodianship of scholarly metadata is essential for institutional integrity, provenance, and fairness. Metadata should be openly accessible to all, but authoritatively curated by institutions.

## References (if applicable)

- [1] Barcelona Declaration on Open Research Information. (2023). Barcelona Declaration on Open Research Information. Available at: <https://barcelona-declaration.org/> (Accessed: 9 January 2026).
- [2] Kunze, J. and Rodgers, R. (2008). The ARK Identifier Scheme. California Digital Library. Available at: <https://escholarship.org/uc/item/9p9863nc>
- [3] Knoth, P., Budko, V., Pavlenko, V. and Cancellieri, M. (2022) OAI Identifiers: Decentralised PIDs for Research Outputs in Repositories. Open Repositories 2022, Denver, USA. Available at: <https://core.ac.uk/resources/oai-identifiers-and-resolver-2022.pdf>

## Community Feedback

The ideas presented in this opinion piece reflect the perspective of **Petr Knoth** and are intended to contribute to an ongoing discussion within the repository and open research infrastructure community about the role of institutions in stewarding scholarly metadata.

We would greatly welcome your reflections, perspectives, or experiences related to these questions. If you would like to share your thoughts, please email **Mologadi Ramushu** at [mologadi.ramushu@open.ac.uk](mailto:mologadi.ramushu@open.ac.uk) with the subject line:

**“Community Feedback\_Authoritative Custodianship of Scholarly Metadata.”**

Selected insights may help inform future discussions and research on this topic.