MORE BRAINS

Improving the open access research information landscape

An analysis of community needs, challenges, and potential next steps

4 June 2025

Photo by Rodolfo Quirós: https://www.pexels.com/photo/selective-focus-photography-cement-2219024/

Executive Summary

This report brings together the outcomes of extensive community discussions on practical steps that should be taken by a range of stakeholders to enable UKRI-funded authors to meet the metadata and technical requirements of the UKRI open access policy. The discussions took place in UKRI's Open Access Policy Technical Requirements Project Group, with their Open Access Stakeholder Forum,[1] and in dedicated community consultation events which took place between September and November 2023.

Consultants from MoreBrains Cooperative facilitated and recorded these sessions; gathered community input by themes, which were identified as part of previous work on policy implementation[2]; and synthesised them to generate a list of concrete actions that should be taken to enable researchers and research organisations to both meet the policy requirements, and to be able to show that they have met them. Community priorities throughout the discussion were clarity, efficiency, and scalability. Wherever possible, proposed activities focus on the widespread adoption of existing standards and good practices, or the extension of existing standards, schemas, or protocols.

The open access publishing landscape is complex, and many stakeholder groups contribute to it. There are established practices, conventions, and workflows in place, which are shaped by the expertise and evolving needs of all those engaged in the publication process. The actions suggested below therefore involve contributions from a range of groups, from national infrastructures to international standards bodies. UKRI has a range of roles to play in delivering the recommendations below, including direct interventions, convening and facilitating discussions, additional investigations, and making recommendations for updates to global standards and services.

It would not be appropriate for every action recommended to be led by UKRI. Further work will be required across stakeholders to agree responsibilities and develop concrete activities to support implementation of the recommendations. The recommendations in this report, in line with the goals for the policy requirements that they address, are intended to improve the functioning and interoperability of publishing and repository services and, therefore, to benefit research more widely. UKRI sees this work as a catalyst for change to help all actors improve their implementation of standards and best practice. Meeting the requirements of UKRI's open access policy is therefore just one benefit among many of the work suggested here. Improved processes will increase operational efficiency in publishing companies and reduce bureaucratic burden in universities, for example.

The shared goal in all these proposed endeavours is to improve access to published research. Improving the completeness of descriptive metadata, consistent standards adoption, the use of persistent identifiers, reliable statements on licensing or preservation, and more alignment of policies are all outcomes that are recognised by the policy, and by the community, as contributing

[1] https://www.ukri.org/publications/open-access-policy-stakeholder-forum/

[2] J. Brown, P. Jones, A. Meadows, F. Murphy, and P. Knoth, 'Metadata to support the UKRI Open Access Policy: Landscape and community readiness analysis'. Zenodo, Nov. 28, 2022. doi: 10.5281/zenodo.7386901

to that goal. The activities proposed below will help to deliver these priorities and will benefit the publishing and research ecosystems alike.

There are five key recommendations, as follows.

1. **UKRI should begin registering Grant IDs because doing so will enable the unique identification of funded awards and robust linking to researchers' research activity in ways that traditional grant IDs do not**
2. **Conduct a range of targeted outreach and communication programmes in collaboration with relevant stakeholders**
3. **Direct publishers, publishing technology vendors, and repository communities to existing technical resources**
4. **Conduct research into barriers to compliance faced by the publishing and repository communities**
5. **Support the technical development of specific plugins to assist repositories**

In addition, the consultations highlighted two areas that are of particular importance to the repository community:

- Guidance on what criteria should be used to select persistent identifiers (PIDs) for content held in repositories, and for other research administrative entities (such as projects, etc.)
- Clarity on what constitutes a 'suitable repository' from a technical standpoint

The work leading up to this report has been largely diagnostic. Our analyses of sector readiness to meet technical requirements of the UKRI open access policy have exposed inconsistent applications of standards and best practice, with many gaps. Filling these gaps will require action from all stakeholders in the research lifecycle, including UKRI. A critical step will be to bring these stakeholders together to identify dependencies, resource constraints, and priorities. It will then be possible to map out a shared programme of improvements with benefits that go far beyond open access.

Overview of actions required to deliver the recommendations

Each of the components of the recommendations is categorised into one of three types, as indicated by one of the three following icons:


Technical - Changes required to the research information infrastructure, new tools or workflows


Education / Outreach - Documentation, best practice, community development


Research - Areas where change needs to happen, but more research is needed to better understand community needs and how to achieve that change

Each of the actions recommended is also broken down by the timeframe it is most feasible or appropriate to expect action to be taken in. Note that where there are dependencies on other actions, seemingly high priority interventions may be pushed back to later phases.

Short-term (one to three years) actions recommended are:

Category	Action	Relevant recommendation
	UKRI should lead by example and roll out Grant DOIs for all relevant funding programmes	1
	Educational resources should be created. There is a need to engage with publisher and publishing technology vendors on the benefits of improving metadata reach through interoperability with PID systems, working with relevant stakeholders like PID-supporting organisations where possible	2
	Clarity on the types of PIDs that are required and/or recommended for both publishers and repositories is needed	2
	Use appropriate communication channels to articulate the benefits of CC-BY licences, thereby addressing community questions and concerns about this	2
	The publishing community should be directed to appropriate educational resources on metadata best practices and technical information about relevant schema and Application Programming Interfaces (APIs)	3
	The publishing community should be directed to appropriate resources on standard XML publication tagging suites	3
	The publishing and repository communities should be directed to resources on metadata standards for describing licensing information	3
	Publishers should be directed to appropriate technical information about long-term preservation	3
	The publishing and repository communities should be directed to appropriate resources on open citation standards and practices	3
	A PID landscape and gap analysis is needed to understand current approaches to PIDs among repositories. Identify pathways to consensus and enable support for multiple approaches where appropriate	4
	Research to identify efficient and scalable options to expand ORCID authentication across the ecosystem is needed	4
	A sector-wide discussion is needed with respect to the financial burden on smaller publishers created by DOI registration. Levels of economic exclusion should be considered and, if needed,	4

mechanisms for support explored

Medium-term (three to five years) actions recommended are:

Category	Action	Relevant recommendation
	The importance of PIDs across the research lifecycle in the UK must be emphasised. This will require sector-wide effort in collaboration with key PID-supporting organisations in the UK and globally, Engagement efforts should include blog posts, industry publications, conference talks, and social media to raise awareness of the benefits that widespread PID adoption brings	2
	A repository community conversation is needed about best practices for PID usage and metadata creation	2
	Messaging and guidance around ORCID authentication mechanisms as they relate to the policy are needed, including the relevance of trust markers, ORCID OAuth, and integrations	2
	Documentation on definitions and best practice in relation to UKRI's policy on self-archiving is needed	2
	Clear and concise guidance on acceptable long-term preservation locations and providers is needed	2
	Many relevant technical standards exist, as provided by various sector infrastructure providers. Communities should be directed to those resources via social media, conferences etc.	3
	Funders other than UKRI should be encouraged to adopt Grant IDs	4
	Research into potential approaches to improved routing of metadata between publisher versions of record and repository copies of outputs should be explored, with a view to improving the quality and completeness of article-level metadata in repositories	4
	Discussions are needed with publishers, publishing associations, and open citation advocates on the needs for best practice and progress in the area of open citations	4
	Exploratory work on the feasibility of machine-readable self-archiving policies is needed, possibly in collaboration with relevant standards bodies	4
	Barriers and opportunities for machine-readable open access status metadata in repositories should be investigated	4
	In the medium to long term (recognising that this work has dependencies on activities recommended above), a series of repository plug-ins need to be developed, to ensure that all	5

appropriate metadata (as identified in MoreBrains' previous report[3]) can be supported, in particular, the integration of PIDs including ORCID, DataCite, Crossref, and eventually RAiD

Long-term (five plus years) actions recommended are:

Category	Action	Relevant recommendation
	Community engagement around openDOAR is required. Use cases should be described and promoted including the benefits of repository registration. CRIS technology providers should be engaged with to explain registration requirements and encourage adoption	2
	There is a need to raise awareness of the value of Sherpa services, and their benefits to other stakeholders including the publishing and repository communities.	2
	With respect to self-archiving (Authors Accepted Manuscript deposit), both repositories and publisher platforms should be assisted in coordinating across UKRI, other funders, institutional, and publisher policies	2
	There is a need to raise awareness about OpenCitations, and its benefits to other stakeholders including the publishing and repository communities	2
	Identification of barriers to compliance with the long-term preservation requirements of the policy is needed. More detailed guidance on long-term preservation will be required	4
	Options for a workflow to computationally test whether repositories are registered in OpenDOAR in an automated way should be explored	4

[3] J. Brown, P. Jones, A. Meadows, F. Murphy and P. Knoth, 'Metadata to support the UKRI Open Access Policy: Landscape and community readiness analysis'. Zenodo, Nov. 28, 2022. doi: 10.5281/zenodo.7386901.

Table of Contents

- **Version 1 submitted to UKRI May 2024**
- **Revised version submitted to UKRI July 8, 2024**
- **Second revised version submitted to UKRI October 23, 2024**
- **Third revised version submitted to UKRI February 27, 2025**
- **Fourth revised version submitted to UKRI March 4, 2025**
- **Final version submitted to UKRI June 4, 2025**

1 Background

The UKRI open access policy for research articles sets out technical standards requirements for journals, publishing platforms, and repositories, covering a range of factors from preservation to article-level metadata. Technical standards support full and immediate open access by ensuring research outputs are findable, accessible, interoperable, and reusable. They can also reduce the burden associated with monitoring and managing research. In 2022, UKRI commissioned a report[4], conducted by MoreBrains Cooperative, on repository, landscape, and community readiness for the technical requirements of the UKRI open access policy[5]. As part of that work, MoreBrains conducted an analysis of the policy text to identify specific pieces of information that would need to be made available in the form of metadata, in order to comply with one of the two routes to compliance with the policy.

The MoreBrains team examined various relevant schema and identified gaps where necessary metadata could not be adequately encoded. Where possible, we then assessed the current levels of metadata completeness for publisher and repository sources. We also found a need to improve levels of schema consistency and standards adoption, particularly among repositories. Our analysis showed that, overall, the landscape currently does not fully support implementation of the requirements set out in the policy — a phased approach to full adoption is necessary. The report made several recommendations for how this can be achieved, which this work builds on.

To support implementation of their open access policy, UKRI then commissioned MoreBrains to help consult with the repository and publishing communities (including relevant vendors) on their current state of readiness to meet policy requirements, including identifying potential next steps to open access policy compatibility that can guide UKRI's work and that of other stakeholders. The team engaged widely with stakeholders on the project, including establishing a project group to support its development, and conducting a series of community engagement activities covering 17 specific areas of interest with respect to the policy.

This report is structured to mirror the structure of the policy, with an emphasis on Annex 2 of the open access policy, in which seven technical requirements (lettered A-G) are provided for journals and publishing platforms, and five (lettered A-E) for repositories. There is a dedicated section (2.1) in the findings that focuses on the technical implications of the main body of the policy. The discussions of the requirements discussed there either encapsulate community feedback, or refer forward to the relevant technical requirement from Annex 2 if the implications are more fully addressed there.

The report includes an actionable list of five high-level recommendations for UKRI and other stakeholders to take forward, which have been broken down into a granular list of specific action points. It also outlines two potential pieces of follow-on work, which strongly emerged from the engagement exercises. Further work will be required across stakeholders to agree responsibilities and develop concrete activities to support implementation of the recommendations.

---

[4] [1]J. Brown, P. Jones, A. Meadows, F. Murphyand P. Knoth, 'Metadata to support the UKRI Open Access Policy: Landscape and community readiness analysis'. Zenodo, Nov. 28, 2022. doi: 10.5281/zenodo.7386901.

[5] https://www.ukri.org/publications/ukri-open-access-policy/

2 Findings

Building on the analysis of the metadata landscape in our previous report for UKRI, we worked with UKRI's Technical requirements Project Group (TRPG) (see section 5.1) to analyse specific challenges, pathways to improved metadata availability, and open questions. This work continued through much of the first half of 2023, and culminated in a detailed programme of community engagement centred on specific aspects of the technical requirements (such as OpenDOAR registration, content preservation, and the collection of ORCID IDs for all UKRI-funded authors). The design and delivery of this engagement programme is set out in section 5.2 below, and the meeting accounts are recorded in Appendix A.

After conducting 15 meetings, both virtual and in person, with 145 total attendees (many of whom attended more than one meeting), we were presented with a rich landscape of social and technical challenges to be addressed, practical suggestions for scalable and efficient improvements to the metadata landscape, and pointers to areas where further investigations would be useful.

In this section, we set out key lessons drawn from the analysis and engagement programme, including actions that could be taken by a range of stakeholders. UKRI could undertake a subset of these actions, and could participate in others, while some are community actions or international work that UKRI can encourage, but not deliver directly. Our recommendations (see section 3) deal with the subset where UKRI may have a concrete role.

Suggestions for improvements to metadata quality and availability broadly fit into three categories:

1. Technical: *Changes required to the research information infrastructure, new tools, or workflows*
2. Education/Outreach: *Documentation, best practice, community development*
3. Research: *Areas where change needs to happen, but more research is needed to better understand community needs and how to achieve change*

We have structured our findings by mapping them to sections of the policy: the general policy requirements (set out in section 2.1), or for each of the two routes to compliance: immediate availability of the Version of Record (VoR) in a journal or publishing platform (set out section 2.2); or immediate availability of the Author's Accepted Manuscript (AAM) in a suitable institutional or subject repository (set out in section 2.3). In either case, the article must be made available under a CC-BY licence (with CC-BY-ND permitted by exception).

Our account of the possible pathways forward that emerged across the engagement process is set out here as context for the recommendations in 3 below. Where community members indicated that actions should be a high priority, or where dependencies were identified, this is included as additional context to inform the recommendations.

2.1 Overall policy requirements

Open licences

For the purposes of this analysis, the requirement for a CC-BY licence or, by exception, an Open Government Licence (OGL) or CC-BY-ND licence, are covered in sections 2.2 and 2.3 in the

sections on machine-readable licences. We focussed on how licences can be expressed and recorded, and this applies equally to all permissible licences under the policy.

Version of Record versus Author's Accepted Manuscript

Section 6(a) of the policy states that, for Route 2, authors must "publish the research article in a subscription journal and deposit the Author's Accepted Manuscript (or Version of Record, where the publisher permits) in an institutional or subject repository at the time of final publication". Aspects of the discussion from this investigation are covered in the sections on persistent identifiers for articles, below in sections 2.2 and 2.3, as the policy requires the PID for the VoR to be included in the repository metadata record.

Feedback was that **it would be helpful for publishers to pursue earlier registration of DOIs for pending publications**, which would enable repositories to make connections earlier in the article lifecycle. **Publishers should also collect AAM PIDs wherever available. This could potentially be achieved via the use of the NISO Manuscript Exchange Common Approach (MECA) Recommended Practice**[6], or other approaches, to automate information exchange between publishers and repositories. MECA enables metadata that has already been entered by the researcher into a repository or CRIS to be automatically transferred into publisher submission systems. The reverse is also possible. **More consistent use of PIDs, and more consistent indicators of which PID is linked to which version in repository metadata, would be beneficial**. Finally, it would be helpful for r**epositories to make better use of the open metadata sources currently available** (such as Crossref, ORCID, and CORE) to fill gaps in metadata and improve consistency across systems.

Date of repository deposit

In order to support the overall aim of the policy, section 6(b) states that "a publisher-requested delay or 'embargo period' between publication of the Version of Record and open access of the deposited version is not permitted". For this requirement to be monitored, **the date of repository deposit needs to be consistently recorded**. At present, this data can be overwritten by subsequent updates, such as a repository platform migration, or changes made in a local CRIS. This makes the data unreliable. **The variability in approaches makes this a surprisingly challenging requirement to address efficiently**. The community felt that **UKRI and other relevant stakeholders should work with repository managers and developers to define the problem space**. This could be followed by **clear statements of best practice, with definitions and specifications** that repository platforms could be measured against.

Data Access Statement

Section 15 of the policy states that "UKRI requires in-scope research articles to include a Data Access Statement, even where there are no data associated with the article or the data are inaccessible". While practice is very diverse in this space, the community felt that **UKRI should continue to evolve its recommendations for this and, wherever possible, align with other funder requirements**. Ideally, this would result in a clear specification of what a compliant DAS would include and how it would be formatted, with clarity on any exceptions. It would be helpful to **steer**

[6] https://meca.niso.org/

authors and publishers toward including DASs as a 'top level' section in articles, to improve machine readability and discoverability. **Repositories, in particular, would benefit from clear guidance on the technical implementations necessary** beyond, for example, the existing support in RIOXX 3.0 for links to related datasets.

2.2 Route one findings: journals and publishing platforms

Requirement A: Persistent Identifiers (PIDs) for articles

The policy states that "Persistent Identifiers (PIDs) for articles must be implemented according to internationally recognised standards. Examples of international standards include Digital Object Identifiers (DOI), Uniform Resource Name (URN) or Handle".

While adoption of PIDs (predominantly DOIs) for articles is very widespread, there is less awareness that much of the value of these PIDs comes from the quality and completeness of their associated metadata.

Suggested pathways to improvement tended to focus on education and outreach. **UKRI and other funders need to articulate the importance of PIDs across the research lifecycle**, especially to smaller publishers/community initiatives. Directing smaller publishers to existing resources (Crossref, etc.) on the broader importance of PIDs and metadata, and why they're increasingly necessary, would be of value. As part of this process, a community goal should be to **educate publishers on the value and reach of their metadata**.

Note that **small publishers may not have the resources to pay for DOIs** or the capacity to engage with metadata improvement.

Requirement B: Article-level metadata

The policy states that "article-level metadata must be used according to a defined application profile that supports the UKRI open access policy and is available, if possible, via a Creative Commons public domain dedication (CC0). The metadata standard must adhere to international best practice such as the Crossref schema and OpenAIRE guidelines".

Crossref has produced numerous guides to metadata, which articulate their value very well. It would therefore be useful to **direct publishers to appropriate Crossref educational resources on proper metadata standards and workflows**.

Engagement with the Journal Article Tag Suite (JATS) standard[7] is a critical pathway to improving the consistency and coverage of article metadata, as it shapes metadata structure separately and upstream from DOI registration.

Requirement C: Machine-readable licences

The policy states that "machine-readable information on the open access status and the licence must be embedded in the article metadata in a standard non-proprietary format".

[7] https://jats.nlm.nih.gov

The existing NISO Recommended Practice on Access & License Indicators (ALI) meets this requirement and it should be expressed as a Uniform Resource Identifier[8] (URI)[9]. **Work to raise awareness of the ALI standard could be reinforced by making the rationale for UKRI's preference for CC-BY explicit in communications and advocacy efforts**.

More research into how machine-readable licensing workflows could be developed, and engagement with the evolution of the ALI standard would help to further incentivise wider adoption.

Requirement D: Long-term preservation

The policy states that "long-term preservation must be supported via a robust preservation programme such as CLOCKSS, Portico or an equivalent".

The **Keepers Registry[10] would be a valuable partner in implementing and monitoring this requirement.** Working with the Registry to develop more detailed guidance or requirements would be of benefit, as would work to research digital preservation approaches in collaboration with exemplary practitioners such as Europe PubMed Central.

In addition, education and outreach should **focus on compiling clear and concise guidance, particularly on acceptable preservation locations and providers**, and it should **direct publishers/service providers to resources such as the Open Archival Information System (OAIS) for technical information[11].**

Requirement E: Citation data

The policy states that "openly accessible data on citations must be made available according to the standards set out by the Initiative for Open Citations (I4OC)".

Since the policy was published, I4OC has been wound up, so **OpenCitations[12], alongside Crossref, should be used as the guide for this requirement**. This work could also be informed by outputs from the Research Data Alliance, including their work on complex citations[13].

In terms of education and outreach, it would be beneficial for UKRI and other stakeholders to **facilitate discussions between advocates, publishers, publishing associations, and OpenCitations**. This work could be reinforced by **participation in coordination, advocacy, and education efforts by OASPA, DOAJ, COPE, and other interested groups**.

Work to **assess opportunities for support and/or collaboration with OpenCitations**, including the ingest of metadata into OpenCitations Meta, and additional **research on the quality of citations including accuracy and presence of DOIs,** would also be valuable.

---

[8] See https://en.wikipedia.org/wiki/Uniform_Resource_Identifier
[9] Following the approach set out at: https://www.niso-sts.org/standard-html/v1-0/niso-sts-1-0/niso-sts-1-0-elem-ali_license_ref.html
[10] https://keepers.issn.org/
[11] http://www.oais.info/
[12] https://opencitations.net/
[13] https://www.rd-alliance.org/groups/complex-citations-working-group

Since DOIs are a key dependency for both this requirement and requirement A (see section 2.1 above), **work to understand affordability and capacity challenges for smaller publishers** could be helpful.

Requirement F: Self-archiving policies

The policy states that "self-archiving policies must be registered in the SHERPA RoMEO database".

Potential education and outreach activities include convening discussions on **standardising language (e.g. around terms for article/manuscript types) and developing documentation on policy information and workflows** (possibly in the form of case studies).

It would be helpful to **work with Jisc to promote and explain the value of Sherpa services to publishers**. Ongoing **work to maintain Sherpa services and to provide assurance of the accuracy and currency of the data** it contains is also needed.

Additional **research may be needed to develop machine-readable policies for self-archiving**, which could be accelerated by engagement with NISO. **Research on how to develop structured formats and workflows for easy reporting,** such as JSON or XML, and **to identify which organisations to collaborate with to implement these,** would accelerate alignment with this policy requirement.

Requirement G: Research management PIDs

The policy states that "common unique PIDs for research management information (for example identifiers for funders or organisations) are strongly encouraged. ORCID, the researcher identifier must be supported".

UKRI could have a major impact in leading by example and issuing Grant DOIs for all awards. The account of the workshop on Grant DOIs in Appendix A6, below, sets out many of the direct advantages to UKRI of this, and the linked presentation and exercises from the workshop outline the core elements of the strategic and business cases for doing so. Wider benefits will include gains for authors, institutions and publishers, as well as helping to foster the practice of using PIDs for grants amongst other funders.

Education and outreach priorities include **collaboration with ORCID on messaging, and guidance on ORCID authentication**. It would also be useful for **UKRI and others to convene stakeholders, including other funders, to support wider Grant ID adoption**. The perceived requirement for ORCID IDs for all UKRI-funded authors has been a source of some concern across communities. Effective implementation of ORCID IDs in line with best practice[14] will require the implementation of changes to workflows and careful communication across stakeholders.

Providing **clarity on which PIDs are recommended and acceptable/appropriate** would also help to focus adoption and integration efforts, enabling UKRI and others to **collaborate with relevant PID providers on messaging to increase adoption**.

[14] As set out, for example, in their certified service provider program: https://info.orcid.org/vendors-and-service-providers/become-an-orcid-certified-service-provider/

Additional research is needed to **identify efficient and scalable options for ORCID authentication,** and transmission of proof of authentication for co-authors' IDs. This could be seeded by Grant DOI metadata, for example, making it easier for authors to collect and verify co-author IDs.

2.3 Route two findings: repositories

Requirement A: PIDs for articles

The policy states that "PIDs for research outputs must be implemented according to international recognised standards. Examples of international standards include DOI, URN or Handle".

A plurality of approaches to PIDs would be best, to enable institutions to meet this requirement in the way that best matches their capacity and available options. Using DOIs for versions other than the VoR would lead to confusion in some institutions, while others may want to leverage existing memberships in, for example, DataCite and are comfortable with the versioning and relationship management available in current DOI metadata schemas. Some prefer a decentralised approach; others want the services provided by a central registry. **There are a range of preferences and priorities in play with respect to PIDs for repository content, and there is no current consensus that points to a one-size fits-all approach**.

Repositories need information to make the choice that is best suited to their business and technology needs. UKRI and others could help develop messaging about what constitutes a suitable repository, building on the existing guidance around the technical requirements.

Informed and evidence-based choices about PIDs could be assisted by providing a PID criteria matrix to assist repositories in making decisions about PIDs. To drive meaningful action on this issue, clarity of purpose and non-negligible investment will be needed, so **an engagement plan should be developed to explain the purpose/importance of the requirements, with a focus on the Vice-Chancellor level,** in order to unlock institutional investment in PIDs.

This senior management focus could be complemented by the **creation of documentation for repository implementers and managers, with a focus on use cases**.

Given the potentially contentious choices to be made, research to complete **a landscape and gap analysis of current repository approaches and use of PIDs,** and to **identify where pathways to consensus need to be found and where multiple approaches need to be supported,** will be valuable. They will enable meaningful movement towards meeting this policy requirement and will help to cut through some of the conflicting priorities and competing interests in play.

Requirement B: Article-level metadata

The policy states that "article-level metadata must be implemented according to a defined application profile that supports the UKRI Open Access Policy and if possible is available via a CC0 public domain dedication. This must include a persistent identifier that resolves to the item landing page displaying metadata and link(s) to the full text of the Author's Accepted Manuscript (where available) or the Version of Record, or both. The metadata standard must adhere to international best practice such as the OpenAIRE guidelines".

This requirement is also served by the corresponding provisions for publishing platforms outlined in section 2.2 above, especially with regard to the provision of **advice on PID selection and the development of repository plug-ins**. **Work to explore ways to improve routing and synchronising AAM and VoR metadata for UK-funded research outputs** will be critical for meeting this requirement.

Highlighting pathways to alignment with this requirement, such as implementing the RIOXX application profile in full and keeping it up-to-date, or fully populating DataCite metadata records when registering a DOI for an AAM, will help organisations to identify a suitable approach.

Repositories must be free to choose the approach that matches their priorities and resources, so clear **guidance on the metadata fields required for policy compliance, and evaluation of which current offerings meet those requirements,** will also be needed.

Requirement C: Machine-readable licences

The policy states that "machine-readable information on the open access status and the licence must be embedded in the metadata in a standard non-proprietary format".

The repository community could be served by the same provisions set out for publishing platforms in section 2.2 above, with some additional information needed to **explore repository-specific barriers and opportunities for machine-readable open access status, in collaboration with the community and repository providers**.

Requirement D: Research management PIDs

The policy states that "common unique PIDs for research management information (for example identifiers for funders or organisations) are strongly encouraged. ORCID, the researcher identifier, must be supported".

Many of the corresponding provisions indicated in section 2.2, requirement G above apply here, with the addition of some repository-specific needs. **Repository plug-ins will need to be developed to enable PIDs and associated metadata to be incorporated effectively**. Aligned with the provisions set out in section requirements A and B above, **convening a community conversation about best practice for PID usage and metadata creation and re-use** supported by the messaging, use cases, and landscape analysis set out above will also be valuable for repository managers.

Requirement E: OpenDOAR registration

The policy states that "the repository must be registered in the Directory of Open Access Repositories (OpenDOAR)".

UKRI and others could help to support compliance with this requirement by **participating in community engagement activities on OpenDOAR,** including use cases, and by **providing clarity around the benefits of being registered in OpenDOAR**, beyond it being considered best practice.

Given that Current Research Information Systems (CRIS) are in use as a repository solution at some organisations, it would be helpful to **communicate registration requirements for CRIS-based solutions**, especially as these are not always recognised as filling the repository role.

Research is needed to **establish a workflow to test whether a repository is registered in OpenDOAR**, and **technical investigation is required to understand the best way to identify the repository**. This could be through PIDs or URLs; in both cases, data availability and accuracy need to be addressed.

3 Recommendations

Our recommendations are based on the findings of the engagement exercises described in section 2, and our discussions with the TPRG, which were informed by our previous analysis of the metadata landscape. Recommendations are presented in priority order, except where dependencies on other actions mean that they have to come later in the implementation sequence. For the purposes of this discussion, short term is defined as within one to three years, medium term is between three and five years, and longer term is five years or more.

Each of the components of the recommendations is categorised into one of three types, as indicated by one of the three following icons:


Technical - Changes required to the research information infrastructure, new tools, or workflows


Education / Outreach - Documentation, best practice, community development


Research - Areas where change needs to happen, but more research is needed to better understand community needs and how to achieve change

Recommendation 1: UKRI to issue Grant IDs


UKRI should lead by example and roll out Grant DOIs for all relevant funding programmes in the immediate or short term. Grant DOIs should be associated with complete metadata, which should be updated to support accurate funding acknowledgements and to trigger policy-aligned workflows, aligned with the good practice set out in Recommendation 2, below. In particular, the metadata should contain related PIDs, including authenticated ORCID IDs for researchers affiliated to UKRI-funded projects.

Recommendation 2: Conduct outreach and communication programmes in collaboration with relevant stakeholders

In the short term:


Educational resources should be created. There's a need to engage with publisher and publishing technology vendors on the benefits of improving metadata reach through interoperability with PID systems, working with relevant stakeholders like PID-supporting organisations where possible


Clarity on the types of PIDs that are required and/or recommended for both publishers and repositories are needed


Use appropriate communication channels to articulate the reasons for UKRI's (and others') preference for CC-BY licences, thereby addressing community questions and concerns about this

In the medium term:

- The importance of PIDs across the research lifecycle in the UK must be emphasised. This would require sector-wide effort and collaboration across key PID-supporting organisations in the UK and globally (e.g., Jisc, Crossref, ORCID, DataCite). Engagement efforts should include blog posts, industry publications, conference talks, and social media to raise awareness of the benefits that widespread PID adoption brings
- A repository community conversation is needed about best practices for PID usage and metadata creation
- Messaging and guidance around ORCID authentication mechanisms is needed, including the relevance of trust markers, ORCID OAuth, and integrations
- Documentation on definitions and best practice in relation to UKRI's policy on self-archiving is needed
- Clear and concise guidance on acceptable long-term preservation locations and providers is needed. Work with Keepers Registry is needed to develop and promote the guidance

In the longer term:

- Community engagement around openDOAR is required. Use cases should be described and promoted and should articulate the benefits of repositories being registered. Clarity on why UKRI need repositories to be registered should be provided. CRIS technology providers should be engaged with, to explain registration requirements and encourage adoption
- There is a need to raise awareness of the value of Sherpa services, UKRI's motivation for supporting them, and the benefits to other stakeholders including the publishing and repository communities. This includes clarifying why UKRI needs publishers to use Sherpa services
- With respect to self-archiving (AAM deposit), both repositories and publisher platforms should be assisted in coordinating across UKRI, institutional, and publisher policies
- There is a need to raise awareness about OpenCitations, UKRI's motivation for supporting them, and the benefits to other stakeholders including the publishing and repository communities, working with OASPA, DOAJ, COPE, and other relevant interest groups

Recommendation 3: Direct publishers, publishing technology vendors, and repository communities to existing technical resources

In the short term:

- The publishing community should be directed to Crossref educational resources on metadata best practices and technical information about their schema and Application Programming Interfaces (APIs) (https://www.crossref.org/documentation/)

- The publishing community should be directed to the Journal Article Tag Suite (JATS) pages hosted on the website of the US National Library of Medicine (NLM) (https://jats.nlm.nih.gov/) for information about best practice in metadata for journal articles (https://www.loc.gov/preservation/digital/formats/fdd/fdd000453.shtml)
- The publishing and repository communities should be directed to documentation of the ALI standard available on the NISO website (https://www.niso.org/standards-committees/ali-revision)
- Publishers should be directed to the reference model for Open Archival Information Systems (OAIS) for technical information about long-term preservation (http://www.oais.info/)
- The publishing and repository communities should be directed to OpenCitations (https://opencitations.net/) and Crossref for technical information about open citations

In the medium term:

- Communities should be directed to relevant technical standards and guidelines via social media, conferences, etc, working with organisations like Crossref, DataCite, ORCID, NISO, National Library of Medicine (with respect to JATS)

Recommendation 4: Conduct research into barriers to compliance faced by the publishing and repository communities

In the short term:

- A PID landscape and gap analysis is needed to understand current approaches to PIDs among repositories. Identify pathways to consensus and enable support of multiple approaches where appropriate
- Research to identify efficient and scalable options to expand ORCID authentication across the ecosystem is needed
- A research and publishing sector-wide discussion is needed with respect to the financial burden on smaller publishers associated with registering DOIs. Levels of economic exclusion should be considered and, if needed, mechanism for support explored

In the medium term:

- Funders other than UKRI should be encouraged to adopt Grant IDs
- Research into potential approaches to improved routing of metadata between the publisher VoR and repository copies of outputs should be explored, with a view to improving the quality and completeness of article-level metadata in repositories
- Discussions are needed with publishers, publishing associations, and open citation advocates on the need for best practice and progress in the area of open citations

- Exploratory work on the feasibility of machine-readable self-archiving policies is needed, possibly in collaboration with NISO
- Barriers and opportunities for machine-readable open access status metadata in repositories should be investigated

In the longer term:

- Identification of barriers to compliance with the long-term preservation requirements of the policy is needed. More detailed guidance on long-term preservation will be required, working with relevant stakeholders including the Keepers Registry, CLOCKSS, and Portico
- Options for a workflow to computationally test whether repositories are registered in OpenDOAR in an automated way should be explored

Recommendation 5: Support the technical development of specific plugins to assist repositories

Recognising that this work has dependencies on activities recommended above, for example, the selection of article-level PIDs may require a plug-in specific to the PID chosen, in the medium to long term a series of repository plug-ins needs to be developed. This would ensure that all appropriate metadata (as identified in MoreBrains' previous report[15]) can be supported, in particular, the integration of PIDs including ORCID, DataCite, Crossref, and eventually RAiD

4 Recommended further work

Two specific follow-on projects emerged from the community engagement events, which were discussed and endorsed by a range of stakeholders during multiple meetings: development of a PID selection matrix; and identifying characteristics of a suitable repository. These projects speak directly to community priorities expressed during the engagement process and would address multiple components of the recommendations above. The process of conducting and promoting these projects will, in and of itself, help to allay stated community concerns and support repository managers and developers in their planning and prioritisation. They should be a short-term priority. These should be complemented by ongoing convening and discussions to bring together relevant stakeholders to address some of the specific challenges that require coordinated activities across multiple communities (highlighted in section 4.3, below).

[15] J. Brown, P. Jones, A. Meadows, F. Murphy and P. Knoth, 'Metadata to support the UKRI Open Access Policy: Landscape and community readiness analysis'. Zenodo, Nov. 28, 2022. doi: 10.5281/zenodo.7386901.

4.1 PID selection matrix

Throughout our discussions, members of the repository community expressed a need to understand which PIDs they should be using and/or recommending. There was a clear desire not only to understand what is needed to comply with the current policy, but also to align with UKRI's direction of travel for the future. With that in mind, we recommend developing a PID selection matrix that could be used by the repository community to help individual repository owners select the appropriate PIDs for both policy compliance and their own specific use cases.

Based on the previously conducted landscape review, the UK PID workflows[16] that were created as part of the Jisc-sponsored national PID strategy work[17], and the results of this engagement study, the matrix would provide a list of PIDs for five priority scholarly entity types (people, institutions, outputs, grants, projects). This would be accompanied by a list of characteristics, including those indicated by the UKRI open access policy, such as community governance, the availability of a public API, etc. Organisations that provide or support the various PIDs to be assessed would be asked to verify information about each characteristic. The resulting matrix of answers would be invaluable for repository owners in selecting appropriate PIDs.

This project would help to align good practice across the repository community (relevant to recommendations 2 and 4), create a framework for decision-making, enable the collection and signposting of documentation and information resources linked to each PID (recommendation 3), and support prioritisation for repository plug-in development (recommendation 5).

This piece of work could be quickly adapted from existing resources, and effort would primarily be required around documentation and promotion.

4.2 Characteristics of a suitable repository

In addition to the need to understand which PIDs they adopt, there are a series of other characteristics, features, and functions needed by repositories in order to fully meet the policy requirements. For example, there is a technical requirement for machine-readable licensing information, which implies that repositories must implement a metadata schema that includes licensing information, such as the NISO ALI standard.

More broadly, for UKRI to be able to monitor compliance with the policy, and for research outputs to be Findable, Accessible, Interoperable, and Reusable (FAIR), metadata interoperability is paramount, i.e., functional API endpoints in addition to a well-formed and standardised metadata schema.

These requirements are technically fairly complex and not necessarily clear to all members of the repository community. A research project is therefore needed to define the characteristics of a repository required in order for it to be compliant with the open access policy. The criteria for suitability would need to be fairly wide-ranging and comprehensive, including detailed information

[16] https://www.morebrains.coop/jisc-pid-workflows/
[17] https://www.jisc.ac.uk/innovation/projects/a-national-persistent-identifier-research-strategy

on specific metadata fields and necessary workflows, to provide adequate guidance for repository owners as well as developers and vendors.

This project would involve desk research to classify requirements and review technical specifications, as well as interviews with key stakeholders to ensure that characteristics are reasonable and achievable. Given the breadth of the stakeholder landscape, the complexity of the intersectional relationship between subject specific, national, and international repositories, and the timelines involved in rolling out new repository developments, we recommend that this project is informed and guided by a stakeholder group (similar in scale to the TRPG) to streamline information-gathering and ensure the best chances of community buy-in.

This project would most directly address recommendation 4. The stakeholder inclusion and consultation/research components also support recommendation 2, provide additional insight into emerging practice to support recommendation 3, and provide a mechanism for convening stakeholders to deliver recommendation 5.

4.3 Immediate next steps

The issues around metadata completeness, consistency of practice, and standards adoption identified through this project require action from many stakeholders across the research and scholarly communications ecosystem. Some of these actions can be taken unilaterally by UKRI, and there are also some areas of activity that it would be appropriate for UKRI to lead.

For the remainder, we recommend that UKRI reconvenes the stakeholders consulted in this report to review the recommended action, to identify priorities and dependencies, and to determine who is best placed to deliver the interventions or improvements. We suggest that the initial focus in on exploring the resourcing and capacity requirements of the highest priority and/or most impactful actions identified.

Continuing to work with stakeholders in this way will enable UKRI to design a programme of actions to deliver meaningful and sustainable improvements to the metadata landscape with benefits far beyond meeting the requirements of this policy.

5 Methodology

5.1 TRPG process

The Technical Requirements Project Group (TRPG) was convened by UKRI to support implementation of their open access policy. During Phase 1 of this work (2021-22), this group met informally, on an 'as needed' basis. With Phase 2 seeking to validate Annex 2, which describes the policy's technical requirements, and to develop a pathway to full adoption, the TRPG was reconvened with a more structured remit and programme of activities.

TRPG membership was made up of technical experts and representatives from institutions subject to the UKRI open access policy, technical experts from relevant domains, systems providers and implementers (including repositories, publishing platforms and other information systems), and relevant community initiatives. Their remit was to be a source of additional expertise and reach to enable the technical requirements of the open access policy to be heard, understood, and implemented by relevant organisations, through providing advice, enabling access to community channels, and reviewing functions for documentation. Many, though not all, TRPG members were also involved in the engagement activities and meetings.

5.2 Engagement plan

5.2.1 Development

The Project team convened a series of meetings with the TRPG, to evaluate what support would be required to enable relevant parties to meet the technical requirements of the policy.

These areas of focus included the 17 technical requirements identified in the 2021 MoreBrains metadata landscape analysis. They included questions around various required metadata fields (e.g. how are they to be accurately populated, can this population process be automated?); PIDs (e.g. are they sufficiently developed, and are they understood and being used correctly by relevant stakeholders?); and authentication requirements (e.g. how can ORCID IDs be reliably supplied for all UK-based authors?).

The focal points of the proposed discussions included: what appropriate support for the development and uptake of tools and plug-ins looks like, and how (and by whom) it can be provided; how to encourage communities to disseminate best practices and standardise processes; how to fine-tune communications with infrastructure providers and international funders to align priorities; and what channels could increase engagement with publisher associations.

As a result of these considerations, and in conjunction with input from the TRPG, the project team designed a matrix of engagement event objectives and stakeholder groups.

5.2.2 Design

The 17 recommendations were first grouped into thematic or stakeholder clusters, which fell into two major categories: those concerning metadata and standards relevant to the repository and/or publishing communities, and those relating to PIDs.

For metadata and standards, we subdivided the issues into those specific only to repositories or to publishers (to avoid requiring publishers to participate in meetings relating to institutional or subject repositories, and vice versa) and cross-cutting issues that required input from both communities.

PIDs are fundamental to many of the requirements of the policy (such as article-level PIDs, or support for ORCID IDs for all UKRI-funded authors), and were the subject of eight of our recommendations. They could be further broken down into issues relating to ORCID IDs, including both collection and authentication processes at scale, mature PIDs (for which fine-tuning or consistency of good practice are a focus), and emerging PIDs (where the focus is adoption and development).

While the topics and focus points for the meetings were being developed, the key stakeholders — the potential attendees and contributors — were also being identified. Sometimes this was straightforward, such as ensuring that specific PID providers were included. In other instances, we approached existing community associations and initiatives[18] and worked with them to ensure that a wide range of community members were included in the events. As well as allowing for a diversity of voices, this also had the effect of reducing the burden of participation for organisations.

To contain costs and maximise inclusion (including international contributions), wherever feasible consultation events were held virtually. For particularly contentious or complex issues, or those requiring lengthier analyses which took longer than is reasonable for a virtual meeting, we organised face-to-face events, both for efficiency and to aid consensus-building.

The three categories of virtual events were:

1. Virtual meeting: intended for smaller groups with a focused topic, these sessions were designed to confirm landscape assessments and agree actions.

2. Virtual roundtable: intended for larger consultative groups, these discussions were based on hearing from all participants about the state of the art, community readiness, or a similarly focused question, with an opportunity to discuss an issue in depth to inform UKRI's next steps.

3. Virtual workshop: intended for larger mixed community groups to explore issues where there was controversy, no clear single path forward, or ambiguity that was otherwise preventing progress on an issue.

A total of 15 events (16 sessions) were organised, as follows:

- Face to face workshops:
    - Article level PIDs
    - Funding acknowledgements
    - Machine-readable licences
    - ORCIDs for UKRI-funded researchers
- Virtual workshops:
    - Grant DOIs (split into two sessions)
    - Versions (Version of Record vs Authors' Accepted Manuscript)

[18] Examples included STM, Publishers Association, and Confederation of Open Access Repositories

- Virtual Roundtables
    - Citation Data
    - Data Availability Statements
    - Date of repository deposit
    - ISSNs
    - PIDs for funders and research-performing organisations
    - PIDs for projects
    - Self-archiving policy
- Virtual meetings
    - OpenDOAR registration
    - Preservation location

5.2.3 Delivery

The events were held between 19 September and 9 November 2023, with the four in-person workshops taking place on 1, 2, and 3 November at Caxton House, London.

The standard approach was for a member of the UKRI project team, usually Sara Ball, to welcome participants and introduce the project at a high level. Members of the MoreBrains team then provided more detail about the specific recommendations and facilitated the discussions. In many cases, other participants also gave presentations on the topic. These were very valuable for the subsequent discussions, helping to establish current state of the art and best practice, and to articulate specific community priorities, concerns, or ambitions.

The virtual sessions were recorded (with appropriate permissions) to ensure an accurate note of the meeting; contemporaneous notes were also taken during all meetings. Some of the interactive sessions in the virtual meetings also produced Menti or Miro outputs, and the in-person activities were structured using flipcharts and Post-It notes.

5.2.4 Synthesis

In order to produce a thoroughly reviewed and validated set of meeting notes, we implemented a standardised workflow. An initial meeting account was drafted by one of the MoreBrains team and then checked by another team member (wherever possible, both team members had been present at the meeting itself). We circulated a clean copy of the notes to the invitees (all session participants, and any invitees who had not been able to attend on the day, plus, for some meetings, additional interested parties as previously agreed). Once their feedback had been incorporated, and each meeting account condensed to a maximum of three to four pages in length, the full set of notes was shared with the TRPG, for final validation. The resulting meeting records were then formally reviewed and accepted by UKRI. All the meeting accounts are included in Appendix A (below).

Once this step was completed and final queries resolved, we used the resulting materials as the basis for the findings and recommendations we have outlined in this document.

Appendix A: Meeting accounts

Note: meeting accounts are listed here alphabetically by meeting topic.

A1: Summary of UKRI In-person Workshop on Article Level PIDs

Recommendation

This full-day in-person workshop focused on MoreBrains' recommendation that UKRI work with the repository community to promote best practices that ensure that specific metadata fields for articles are populated and exposed through agreed mechanisms.

Meeting date

November 1, 2023

Attendees

- Todd Carpenter (NISO)
- Melissa Harrison (EBI/EMBL)
- Mark Heaver (Publishers Association/Taylor & Francis)
- Ginny Hendricks (Crossref)
- Petr Knoth (CORE/The Open University)
- Catriona Maccallum (OASPA/Wiley/Hindawi)
- George MacGregor (UKCoRR/University of Glasgow)
- Oksana Parylo (Frontiers)
- Alok Pendse (Publishers Association/Elsevier)
- Ed Pentz (Crossref)
- Iratxe Puebla (DataCite)
- David Sommer (ALPSP/Kudos)
- Sandra Townsend (eJournal Press/Wiley)
- Nathan Westgarth (Aries Systems/Elsevier)

Facilitators

- Sara Ball (UKRI)
- Josh Brown (MoreBrains)
- Phill Jones (MoreBrains)
- Fiona Murphy (MoreBrains)

Notes

NB: The views summarised are those of participants and do not necessarily reflect the views, priorities and policies of UKRI.

Workshop started with a series of presentations from Sara Ball (UKRI), MoreBrains, Ed Pentz (Crossref), and George Macgregor (UKCoRR) and Petr Knoth (CORE). This was followed by group discussion:

- Crossref have a mechanism for pending publication, which allows DOIs to be registered prior to publication, however, it is not yet widely used
- Metadata records in repositories should be much richer. George MacGreggor suggested that ideally parties other than Crossref members should be able to make assertions to metadata records linked to DOIs to enrich those records That could then feed into repositories
- Catriona MacCallum suggested that POSI principles are a good way to signal support for open metadata
- There is a tension between making metadata open and user controlled, at least for ORCID, where researcher choice about openness is a governing principle. ORCID does not make metadata open or closed, that's up to users and organisations, who must be convinced to make the data open
- In response to the suggestion from CORE for OAI IDs as the primary PID for repository metadata records, Todd Carpenter suggested that the persistence level of OAI-IDs is not strong enough and pointed out the lack of an associated metadata schema. There is disagreement about which technology is best suited for the systemic needs of the research ecosystem. Petr Knoth responded that he regards the fact that OAI IDs can be linked to different metadata schemas, such as Dublin Core and Rioxx, is a design strength rather than a weakness of OAI IDs, and that no existing PID is truly persistent, including DOIs. He argues that all PIDs are at most "persistable", i.e. persistent as long as the underlying infrastructure is persistent. In Petr's view a strong repository infrastructure will guarantee the persistence of OAI IDs. (This relates to the observation in the presentations that persistence is a feature of organisations, not technologies.)
- The challenge of persistent identification of resources and interoperable, complete and timely metadata are associated but separate issues, any chosen solution to either one will affect the solution to the other
- Guidance is required to enable repositories to understand and solve the two interconnected problems

Group exercises

Following the discussion, there were two group exercises:

Defining PIDs

Attendees were split into groups and asked to come up with desirable properties and definitions for PIDs in 15 different areas that might help a repository decide which PID(s) are suitable for them. Overall, a need emerged for a sustainable and equitable business model with a firm continuity plan that precludes the PID from becoming proprietary. Governance should be open and community driven. Technically, there's a need for uniqueness by design with adequate, standardised metadata. Broad adoption is needed, with defined use cases for multiple stakeholders. PIDs with broad adoption already are preferable. In the list below, properties that were considered most important or were mentioned repeatedly are underlined.

- Resilience
    - Business continuity plan & back up plan
    - Fallback agreement with another organisation
    - Backward compatibility
    - Long term funding and adequate resourcing

    - Documentation and knowledge transfer
- Open
    - Accessible and easy to discover
    - Free to consume (especially for researchers)
    - Charges for services (not for metadata)
    - Open code, API, data, standards, governance, policies, etc
- Trustworthy
    - Reliable with good provenance
    - Persistent, sustainable and open
    - Community governed
    - Public set of values and principles
    - Adopts POSI principles
    - Cannot be purchased
- Governance
    - Community governed
    - Community engagement
    - Transparent structure
    - Elected board with broad representation
    - Whistleblower policies
- Interoperable
    - FAIR, CARE & trust principles
    - Standards based
    - Integrated with web standards
    - Consistency
- Descriptive
    - Embedded metadata
    - Minimum metadata schema
    - Updateable
    - Maintained and supported
    - Opaque identifier (ID does not contain metadata, e.g. title, container, etc)
- Unique
    - Uniqueness by design
    - Version management
- Resolvable
    - Mechanisms for addressing problems
    - Updateable
    - Verification mechanism
- Sustainable
    - Transparent business model
    - Organisational commitment and resourcing
    - Environmental impact assessment
    - Business continuity plan & back up plan
- Business model
    - Price transparency
    - Equitable pricing
    - Charges for services (not for metadata)
    - Not for profit

- Adoption
    - Adequately scalable
    - Globally usable / relevant
    - Relevant across disciplines
    - Cross-stakeholder value
    - Critical mass in network
    - Outreach program
    - Training and support resources
- Support
    - Repository plugins
    - Technical support
    - Support for metadata schema
    - Support for low- to middle-income countries
- Fit for purpose
    - Appropriate metadata schema
    - Clear use cases for multiple stakeholders
    - Forward and backward linkability
    - Version and relationship management
- Maintained Infrastructure
    - Secure
    - Reliable
    - Scalable
    - Service level agreements
    - Support users as appropriate
- Updateable
    - Open protocols, source code, data
    - System, schema and metadata all updateable
    - Open culture
    - Expectation of updates

Choosing a PID

The attendees were split into three groups and asked each group to identify considerations for the selection of article-level PIDs in repositories. Key findings were as follows:

- Identifiers and entities
    - Are particular identifiers mandated?
    - How easy / cost effective is it to implement?
- Coverage
    - What exists and is in common use?
    - Does my consortium have a preferred or required option?
    - Do support services or networks exist? (Do I have to go it alone?)
    - How important is international adoption vs local/UK adoption?
- Commercial / governance considerations
    - What is the governance / for-profit / not for profit status of the infrastructure provider?
    - Is there a risk of vendor lock-in?

    - What protections exist to prevent the solution becoming proprietary or closed in the future?
- Internal to the institution
    - What do we need to identify / What types of outputs are in the repository?
    - What are the use-cases for the entities in question? (May be institution-specific)
    - What training needs at the institution are needed for successful implementation?
    - What workflows are needed to support the identifier?
    - What reporting requirements / obligations do I need to support? (e.g. impact assessment)
    - What do existing platforms and products used by the institution support?
    - Does the identifier resolve to the repository or publisher instance of an output? Can I control where it resolves to?

Group discussion and recommendations

The group agreed that a PID criteria matrix to assist repositories in making decisions about PIDs would be useful and aid adoption. This effort should be undertaken by UKRI in collaboration with PID providers and repositories. The approach would be for a set of questions to be developed by a group or task force including PID providers and repositories, which PID providers could then answer for themselves.

UKRI should explain the purpose and goals of the requirement in order to give context to this effort. It is also important to engage at the Vice-Chancellor level, to encourage institutional understanding and buy-in in order to unlock institutional investment and resources to deliver effective and efficient PID integrations.

Potential partners could include: Crossref, DataCite, institutions (potentially via UKCORR), repository providers (e.g. EPrints, DSpace, etc), and repository indexing service providers, specifically, CORE.

Documentation should be provided alongside this resource to help repositories get started and should:

- Set out the purpose and scope of the UKRI OA policy as it relates to PIDs, metadata and technical requirements
- Illustrate use cases by example (e.g. assessment, discoverability)
- Make the relationship clear between PIDs and metadata (i.e. they are related but no the same)
- List both PIDs and schemas, how they relate to each other
- Explain the relative benefits and trade-offs for each option

A2: Summary of UKRI Virtual Round Table on Citation Data

Recommendation

The focus of this virtual round table was MoreBrains' recommendation that UKRI work with Open Citations to develop mechanisms for cross-checking whether a publication's citations have been included in their databases.

Meeting date

27 September 2023

Attendees

- Fred Atherden (ALPSP/eLife)
- Mike Cunningham (Frontiers)
- Rocio Gaudioso (Frontiers)
- Ginny Hendricks (Crossref)
- Matt Hodgkinson (COPE/UKRIO)
- Catriona MacCallum (OASPA/Hindawi/Wiley)
- Leila Moore (Publishers Association/Wiley)
- Richard O'Beirne (Publishers Association/Oxford University Press)
- Oksana Parylo (Frontiers)
- David Shotton (OpenCitations)

Facilitators

- Sara Ball (UKRI)
- Phill Jones (MoreBrains)
- Fiona Murphy (MoreBrains)

Notes

NB: The views summarised are those of participants and do not necessarily reflect the views, priorities and policies of UKRI.

Following this presentation by Sara Ball, Phill Jones, and David Shotton, and after confirming that the scope of this element of the policy is journal articles, the main areas of discussion were:

- Current citation practices are ambiguous and historically there have not been good preservation rates. The problem has not been well understood. However, funders are increasingly interested in this issue
- I4OC recommendations include: publishers to deliver references along with bibliographic metadata to Crossref and elsewhere; libraries to require submission of open references and full bibliographic metadata (including author ORCIDs, author institutions, grant numbers and funder IDs) in all new subscription contracts with publishers; repositories using DOIs should submit references with metadata for their holding to Crossref or DataCite, or, if not using DOIs, submit them directly to OpenCitations (N.B. OpenCitations, for lack of sufficient financial support, is not at present able to receive and ingest such submissions into OpenCitations Meta. Far better that publishers be aided to use DOIs and submit references to Crossref.); universities should collaborate to provide open academic analytic services; funders should directly support open science infrastructures such as OpenCitations
- For publishers and international infrastructure providers there are tensions between UKRI's policy, which is nationally based, and the international scholarly communications ecosystem
- Crossref sees I4OC's first recommendation as reasonable. 16,000 publishers (out of 20,000 members) are depositing references via Crossref. Small journals are often doing it

via OJS, which has a Crossref plug-in. This also means that many non-UK focused titles and publishers are active in this space
- However, there are tensions as not all small publishers can afford to use/mint DOIs. For 'Scholar Publishers' OASPA will cover the cost of membership in Crossref and pay for up to 50 DOIs per annum but does not mandate DOIs or deposition of citations to Crosssref for members for equity reasons
- Furthermore, the low-hanging fruit have already been harvested. The success of I4OC was that publishers didn't have to do much in the way of technical action in order to open the references they had already submitted to Crossref. Further progress will be more difficult
- OpenCitations Meta Identifiers (OMIDs) are used internally within OpenCitations Meta to uniquely identify all citing and cited publications involved in citations indexed by OpenCitations, and to disambiguate and reconcile them. An OMID provides an external persistent identifier (PID) for publications lacking a DOI, PMID, etc
    - OMIDs are used internally by OpenCitations as PIDs for every publication that holds OR LACKS an external PID, and act to disambiguate and reconcile them, and ensure there is a PID assigned to each publication recorded in OpenCitations Meta published without an external PID
- COPE is similar to OASPA in its attitude towards standards. Its current requirements are stringent (it turns away 80% of potential applicants). If it were to require high quality metadata, then it would turn away even more
    - As an aside, COPE moved away from best practices to core practises a few years ago, but is now moving back towards the best practices
- There are also geographical considerations, with Russian and Chinese publications being unlikely to comply
- Some artefacts, such as ancient texts, will never have DOIs, although there are potential workarounds, in this context the use of OMIDs is a relevant example

Possible next steps

Improve levels of reference submission compliance

- UKRI could consider supporting OpenCitations, specifically to permit it to ingest into OpenCitations Meta the bibliographic metadata and associated references for publications lacking DOIs. Much of Jisc's funding from UKRI is via Research England, although a feasibility study would be required
- Large publishers and those using established platforms are already doing this. Need to support smaller, scholar led journals
- Support for digital scholarship education
- Take a disciplinary approach, to account for variations in disciplinary norms
- OASPA could recommend that its members start using OpenCitations for reference submissions, given that there is now a non-DOI pathway using OpenCitations Meta. (N.B. OpenCitations, for lack of sufficient financial support, is not at present able to receive and ingest such submissions into OpenCitations Meta. David Shotton suggested that it may be better that publishers be aided to use DOIs and submit references to Crossref)
- UKRI could support OASPA and DOAJ in outreach to its members
- COPE could include best practice recommendations around machine readability, open references and using DOIs

- Consider a tiered approach/phased mandates to the policy that takes into account the likely capacity of the publisher to comply. This may be differentiated according to publisher size and STEM vs HSS
- There are additional complications with partner journals (generally owned by learned societies and published on their behalf by commercial publishers)
- Subsidise DOI creation for journals that cannot afford them
    - Specifically, could UKRI cover the Crossref costs for Diamond OA Journals?
    - Consider working with a pilot group (especially with small/micro publishers) to use as a testbed towards compliance. OASPA and DOAJ could be partners.
- Encourage DOI use in references themselves
    - Investigate whether DOIs are being stripped out during the production processes (this may be the case for HSS workflows)
    - Encourage use of DOIs in reference styles
- Work with platform vendors to achieve implementation of standard improvements.

Ways for UKRI to monitor compliance

- Crossref API or look-up service for checking what each member is doing with metadata: (https://prep.labs.crossref.org/). Crossref also has a Jupyter notebook that aggregates data
- Data Citations corpus via the DataCite API
- Aggregators such as OA.Report (https://oa.report/) which is partly funded by SPARC

Note that the role of pre-prints/preprint servers and grey literature are currently out of scope.

Useful links and resources

- I4OC recommendations
- Open Citations

A3: Summary of UKRI Virtual Round Table on Data Access/ Availability Statements

Recommendation

This virtual meeting focused on MoreBrains' recommendation that UKRI should coordinate across the community to define good practice and reporting pipelines, and develop standards in machine readable data access/availability statements (DAS); work with other funders and publishing associations to encourage uptake of DAS in publishing workflows; work with RIOXX and Datacite to include data access statements in schemas; and promote integrations into other metadata systems like CRIS, Publications Router, and Crossref.

Meeting date

6 October 2023

Attendees

- Dominique Capostagno (NIHR)
- Todd Carpenter (NISO)
- Melissa Harrison (eLife)
- Anton Heimann (Aries/Elsevier)
- Iain Hrynaszkiewicz (PLOS)
- Daniel Keirs (Publishers Association/IOP Publishing)
- Petr Knoth (CORE)
- Hylke Koers (STM Association)
- Adam Leary (Publishers Association/Oxford University Press)
- Catriona MacCallum (OASPA/Wiley)
- George Macgregor (UKCoRR/University of Glasgow)
- Kirsty Merrett (University of Bristol)
- Lou Peck (ALPSP/The International Bunch)
- Iratxe Puebla (DataCite)
- Nigel Robinson (Clarivate)
- Rebecca Taylor-Grant (Publishers Association/Taylor & Francis)
- Marta Teperek (NWO)

Facilitators

- Sara Ball (UKRI)
- Phill Jones (MoreBrains)
- Fiona Murphy (MoreBrains)

Notes

NB: The views summarised are those of participants and do not necessarily reflect the views, priorities and policies of UKRI.

A series of questions was asked via Menti (so the responses were anonymised):

- What is needed to make journal DASs machine readable as standard?
- What needs to happen to enable repositories to collect and structure this data?
- How would it be best for this data to be made available to UKRI?
- How can UKRI help?

Following this presentation by Sara Ball, Phill Jones, Kirsty Merrett, and Iain Hrynaszkiewicz, and after confirming that the scope of this element of the policy is journal articles, the main areas of discussion were:

- The technical infrastructure that supports institutional repositories isn't suited to support the policy via Route 2, UKRI therefore wants to take a phased approach to implementation
- Machine readable, standard wording and consistent application of rules would make this aspect of the OA policy easier to monitor and measure (and would support higher levels of compliance)

- Many publishers have been working on implementing DASs. There are 1.7M papers in PubMed alone with them. Work started in late 2000s, initially driven by the EPSRC policy, with a number of initiatives since. The rates of DASs in articles are generally trending upwards at a modest pace
- There have been funder initiatives with a number of funders implementing DAS policies. Not all funders do so, with some push-back due to concerns about the quality of DASs and their level of effectiveness in driving open data
- For the publisher/journal route (Route 1), it has been discovered that while policies lead to more DASs, they do not necessarily lead to more data sharing. Quality of DASs, when mandated, is generally poor in terms of how useful they are. There are cost and resourcing implications, as well as a need to harness researcher/author motivations and develop standards
- Consistent messaging from publishers may help. Work is being done through RDA where publishers are starting to have those conversations. (Contact Catriona MacCallum for more information)
- DASs are a mechanism to supply contextual information for the research article and data - encouraging data sharing (which may be better achieved through encouraging data citation) does not have to be the primary motivation for DAS mandates
- It would be helpful for UKRI to clarify its main goal for issuing this policy. It is currently unclear whether they prioritise data sharing overall, an evaluation of the reach of open data, or to create context between research articles and supporting data
- For institutional repositories, creating structured metadata fields for DAS is currently of limited value because so few articles have DAS metadata associated with them and so much data isn't being put into repositories
- There is also a question as to whether the article metadata is the correct place to describe the relationships between entities that can be thought of as individual parts of a multi-part research output or project in an institutional repository
- Publishers have in the past taken an approach of recommending repository lists. This has challenges because there is a long tail of data that isn't part of defined data types like gene sequences and researchers may not know which would be a suitable place to put their data, resulting in the policy being ineffective. Therefore, national and/or global data strategies are required
- A publisher has calculated it takes about 25 minutes of staff person time per paper to deliver their mandatory DAS and data sharing policy (not accounting for author/researcher time)
- There is a lack of understanding of both policies and technical requirements for small and scholarly publishers (and diamond publishers) and no resources with which to build the workflows to implement a policy
- Problems for standardisation include the disconnection between the people who want (because they would benefit from) a standard and the people/companies that have to pay to implement the standard that others would use
- There is a challenge for UKRI to materially influence publishers in a timely fashion - publishers are usually globally based so are not directly subject to national policies across their entire operations. Publishers are concerned about over-burdening submitting authors and will tend to rely on advice from their editorial boards (who also tend to have antipathy towards introducing additional administrative burdens without sound scientific rationale). Given that some of the issues are around cost - of developing, implementing, etc - UKRI

can help by issuing a toolkit for publishers, working with other funders to encourage aligned policies to emerge more widely, publish FAQs and other information for researchers, and clarify when the policy DOESN'T apply
- What needs to happen to enable repositories to collect and structure this data?

Existing initiatives and possible options to explore

- Petr Knoth and Melissa Harrison are starting a new project - SoFAIR. It involves automatically extracting dataset mentions. It would be helpful to align thinking here with the Make Data Count Global Data Citation Corpus
- There was an appetite at RDA Salzburg to have that conversation. (Catriona MacCallum is collecting names)
- FAIR Sharing - It's important to register the policies as they are realised
- Encourage better coordination regarding the technical expectations of the repositories? Is CoreTrustSeal relevant here?
- UKRI to align with other funders. It would be helpful if funders were each saying the same things to their researchers.
- There is work ongoing at Crossref to better capture and tag data into the Crossref database.
- A step towards machine readability would be to make DAS a "top level" section in an article
- Europe PMC mines data associated with literature by finding accessions cited within text and also getting links from data repositories where not found within text (i.e. author links their paper to the data but not vice versa)
- Repositories are (in most cases) checking for DASs within papers, so the collection isn't the difficult part. It is arresting the continuing preponderance of papers without a functioning DAS as well as supporting the semantic encoding of DASs - JATS4R is likely best for now.

How can UKRI help?

- Clarify what constitutes a compliant DAS, including what information it requires, and in what format. See for instance, this recently released STM, DataCite, and Crossref statement
- What are the technical requirements for repositories? Would it be sufficient to have a link between the ms and the dataset pointing to where the data is available. Can that be interpreted as a DAS? If so then RIOXX 3 provides. If not, then not.
- Clarify how service providers (e.g. publishers, platforms will benefit from the adoption of the same DAS standard that the administration and funders are seeking to implement

Links and resources

- The Impact on Authors and Editors of Introducing Data Availability Statements at Nature Journals
- Introducing a data availability policy for journals at IOP Publishing: Measuring the impact on authors and editorial teams
- In relation to metadata standards for data submission CEDAR is an important project to look at in terms of creating metadata templates for e.g. different disciplines: https://metadatacenter.org/
- This is the filter one publishing representative uses on PMC to get papers with / probably with DAS https://www.ncbi.nlm.nih.gov/pmc/?term=has+data+avail%5Bfilter%5D
- UKRN Open Research Indicators pilot

- The latest version of RIOXX provides the opportunity for describing datasets and links between manuscripts and datasets. Rioxx delivers those relational links to datasets, from which some inferences about data access statements can be made; but nothing on the nature of that statement or how the data relates to the research in an accompanying paper
- The basic infrastructure exists in JATS and has since 2017. That basic information could be built upon in other structures. There is standardisation work already done for publishers: https://jats4r.org/data-availability-statements/
- The JATS data model has some semantics. Here is the guidance on the use of DAS in JATS, https://jats4r.org/data-availability-statements/
- STM resources for publishers
- McIntosh, Leslie D.; Sumner, Josh; Vitale, Cynthia (2021). Transparently Reported Research: An analysis of Wellcome-funding publications in 2016 and 2019. Wellcome Trust. Online resource. https://doi.org/10.6084/m9.figshare.13810220.v1
- https://datascience.codata.org/articles/10.5334/dsj-2020-005

A4: Summary of UKRI Virtual Round Table on Date of Deposit

Recommendation

This virtual round table focused on MoreBrains' recommendation for the development and adoption of best practice approaches that ensure accurate recording of the date of deposition, which may require workflow changes at institutional repositories; technical fixes to various repository softwares and other systems that repositories integrate with; and community agreement on best practice.

Meeting date

October 10, 2023

Attendees

- Justin Bradley (Eprints/Southampton University)
- Matt Buys (DataCite)
- Mark Hahnel (Figshare)
- Nicolaj Lock (PURE)
- George Macgregor (Rioxx/UKCoRR/University of Glasgow)
- Agustina Martinez-Garcia (DSpace/Cambridge University)

Facilitators

- Sara Ball (UKRI)
- Phill Jones (MoreBrains)
- Fiona Murphy (MoreBrains)

Notes

- NB: The views summarised are those of participants and do not necessarily reflect the views, priorities and policies of UKRI.
- Following this presentation by Sara Ball and Phill Jones, and after confirming that the scope of this element of the policy is journal articles, the main areas of discussion were:
- RIOXX have been engaging with repository managers and learned that decisions about what systems are commissioned, how they are implemented and integrated, and information priorities are often made by others, such as IT or the research office, who are less aware of issues around interoperability and metadata
- There is consensus among repository managers around required dates. However, there is disagreement about how dates are interpreted (e.g. what does date of deposit mean) and a lack of uniformity about how the dates are currently being exposed, and there is likely divergence across schemas
- From the MoreBrains report - less than half of UK repositories have adopted the RIOXX standard. So a little over a half are working on what is possibly a Dublin Core schema. Many say they're OpenAIRE compliant, but they may not be when the data are examined closely. So there are vast differences in levels of metadata completeness which are impossible to measure in the repositories that have not not adopted RIOXX
- The MoreBrains report also recommended that repositories be supported to do one of two things: either adopt the latest edition of RIOXX properly or use DataCite and put the metadata into that and send it to DC for UKRI to interrogate via the API
    - Not all repositories in the UK work with DataCite and there is reluctance from DataCite to participate in any sort of mandate. There could be more federation around technical solutions
    - For context, there are 166 repositories in the British Library DataCite consortium
- If metadata are not exposed consistently, they cannot be aggregated. Dublin Core is so flexible you can be compliant without the metadata being comparable
- The version most RIOXX repositories are currently using won't expose the data. Only the most recent drafts do, and no one is using this yet
- Many of the systems (e.g., PURE, DSpace, or Eprints) can be configured to capture the relevant dates automatically
- Dates can therefore be inferred automatically, but it currently isn't being systematically and consistently recorded in the correct place in schemas, or exposed properly
- It was proposed that relevant dates may not necessarily have to be stored in the repository metadata records but could be harvested by UKRI from Crossref and DataCite records using content negotiation. This mechanism is closer to the source of truth of the information and therefore more reliable
- The Publications Router provides a lot of published versions with high quality metadata as well as the Version of Record full text. Most repositories would have at least the AAM versions. Furthermore, it is uncommon for authors' work to be held elsewhere than in their own institution's repository - except in the case of suitable disciplinary repositories
    - It is not yet clear what constitutes a suitable disciplinary repository for UKRI There is feedback that guidance on this would be useful and that currently researchers are being pointed towards re3data or FAIRsharing.org

    - How might a researcher check their options and ensure that both their institution and UKRI are aware of where their work is in these cases?
        - Collating where their researchers are depositing associated research data is a common institutional challenge. Some institutions are currently querying DataCite by affiliation
    - Possible inclusions here would be PMC or EuropePMC or arXiv?
    - Does there need to be a more exhaustive list?
    - Are aggregators potentially part of the problem or part of the solution? Would they have a place in building the consensus or endorsement of schemas?
- There was a sense that UK repositories have not engaged proactively with OpenAIRE

Possible next steps

- UKRI to do a full landscape analysis on repository schemas, conduct a gap analysis and develop a firm technology strategy for aggregating needed metadata
- Work towards API endpoint consistency. A relevant example is the generalist repository ecosystem initiative (GREI), sponsored by the NIH.
- Support aggregation approaches -, CORE is already actively involved, if this emerges as a feasible approach, then CORE could retain an intermediary role
- Alternative mechanisms for inferring and comparing publishing with date of deposit dates should be reviewed, including the use of aggregators and content negotiation
- Work towards consistency in the way that the data are presented, so that RIOXX, CORE, etc., all display compliant end points?

How UKRI can help

- Documentation to create clarity about the specific dates needed and how they should be defined (The latest version of Rioxx provides a number of specific date properties, with appropriate definitions.)
- Is further evidence needed, in which case, could UKRI commission further evidence gathering?
- An educational piece by UKRI, working with UKCoRR and others, could help with establishing and acculturating best practices for repositories using (the new version of) RIOXX. It could also include information on why interoperability and metadata are important
- Work with repository managers and administrators so UKRI can create tangible incentives to improve practices (via policy?). There could be pathways to improvement where repository managers don't see direct benefit to themselves, but where they can help support practices to improve the information flow more generally
    - Provide endorsement for suitable schemas and support repositories with completing implementations
- A standard definition of the date of deposit, endorsed by UKRI, would minimise the current situation where different institutions and platforms interpret it differently
- Given that metadata quality will likely be poor unless those inputting the information (and those advising and supporting them) are incentivised to provide good quality information, it would be helpful to develop downstream use cases
- UKRI could explore funding the development of technical plug-ins

- Could UKRI help create a more explicit set of prescriptive rules? The community does need to develop its own solutions, but UKRI can help by convening and supporting. What level of compliance requirements is optimal?
    - UKRI will have to make a decision at some point about how much enforcement is optimal
- UKRI could set up a committee and ask for input around schemas, run a pilot, and ascertain the common properties and address gaps
- UKRI could do a landscape analysis on different approaches and which repositories are using DataCite

Helpful links and resources

- https://www.slideshare.net/martinklein0815/comparing-the-performance-of-oaipmh-with-resourcesync
- Rioxx v 3.0 draft: deposit_date: The deposit_date attribute (if present) takes the date on which this resource was first deposited, irrespective of any relevant embargoes or dark archiving, and irrespective of any subsequent file replacement(s). It is anticipated that in some circumstances the deposit_date will be captured and exposed in repository metadata when the resource described is under temporary embargo or temporary dark archiving. If included, this attribute's value MUST be encoded according to the W3CDTF (a profile of ISO 8601) which typically follows the following format: YYYY-MM-DD

A5: Summary of UKRI In-Person Workshop on Funding Acknowledgements

Recommendation

This in-person workshop focused on MoreBrains recommendation that UKRI should work with infrastructure providers, research managers, publishing associations, and publisher workflow systems vendors to develop consensus on a pathway to implementing the inclusion of funding acknowledgements in research articles.

Meeting date

2 November 2023

Attendees

- Imogen Batt (Publishers Association/Springer Nature)
- Francesca Buckland (Clarivate)
- Todd Carpenter (NISO)
- Eleanor Dumbill (CoSector)
- Ginny Hendricks (Crossref)
- Kelly Hetherington (UKCoRR/Durham University)
- Daniel Keirs (Publishers Association/IOPP)
- Catriona MacCallum (OASPA/Wiley)
- Agustina Martinez-Garcia (UKCoRR/Cambridge University)

- Fabienne Michaud (Crossref)
- Iratxe Puebla (DataCite/Make Data Count)
- Aditya Sehgal (Publishers Association/Elsevier)
- Sandra Townsend (eJournalPress/Wiley)
- Nathan Westgarth (Aries/Elsevier)

Facilitators

- Sara Ball (UKRI)
- Josh Brown (MoreBrains)
- Phill Jones (MoreBrains)
- Fiona Murphy (MoreBrains)

Notes

NB: The views summarised are those of participants and do not necessarily reflect the views, priorities and policies of UKRI.

Following this presentation by Sara Ball, Fiona Murphy, Nathan Westgarth, Daniel Keirs, Agustina Martinez-Garcia, and Francesca Buckland, a verbal update from Josh Brown on the Grant PIDs workshop conducted in October, and confirmation that the scope of this element of the policy is journal articles, the main areas of discussion were:

- Persistent identifiers for grants are the mechanism for the policy to use. However, their use cannot yet be mandated as they are not in widespread use
- Currently, it is sufficient to include text about the funding. However, text string acknowledgements hidden in pdfs makes for manual process. Reporting from researchers is often late and partial
- UKRI core-funded researchers also need to acknowledge their funding
- It was noted that if the research object is moved without the information, that doesn't contravene the policy (although maybe it should)
- Adoption of Grant IDs is not (yet) widespread, but usage is growing, and accelerating. There is a need to build sector-wide consensus
- The implementation of Grant IDs would help inform publishers of funders' policy requirements
- Given that funders are able to register Grant IDs at the point of the award, their implementation has the potential to improve the information exchange landscape at an early point in the research cycle (rather than when the outputs are published)
- The use of ORCIDs in grant systems has the potential to generate funding acknowledgement from the existing metadata
- To improve the landscape, schemas need further development and opportunities to maximise interactions between PIDs need to be taken

Questions arising from the discussion

- What are the other pieces of information that could be used to trigger workflow/time/efficiency/information savings? Who needs to do it? What can we do to encourage more uptake?

- Should the acknowledgements be embedded in the research object itself or in the repository metadata record?
- What role could author language editing services and other third-party systems play?
- How will the move to ROR work in practice? Hopefully it will not reduce interactions

Group exercises

Part of the session consisted of a series of group exercises, which focused on identifying opportunities for and barriers against the efficient collection of funding acknowledgments. A summary of the exercises and outcomes is below.

Exercises: How can we collect funding acknowledgements more efficiently? Then, how can we make this a reality?

The brief is to:

- Start with the ideal workflow that would be found in a perfect world
- Consider the perspectives of all relevant stakeholders
- Begin at the grant award stage
- Capture information as early in the process as possible
- Prioritise reuse and reduce rekeying
- Use existing infrastructure (but extra metadata field suggestions are ok)
- Consider who is going to be doing the actual work

Grant award stage - sticky notes:

- ROR, Ringgold, ORCID, Grant DOI
- Project metadata RAID
- PID for use of equipment and facilities
- Unique funder PIDs globally
- Guidance for authors
- Funder tells author their grant DOI and to share it with other systems (preprints, data repositories, publishers etc)
- Funder to be in Open Funder Registry / ROR
- Funder register Crossref Grant
- Funders use ORCID for researchers
- Consistency of funder requirements
- Cross funder/government policy alignment
- EU data transfer outside the EU
- Funders need to get grant management systems that support PIDs

Research in progress stage - sticky notes:

- PIDs for data management objects
- Link to grant funder via a PID
- Research facilities (e.g. STFC)
- See CSIRO pilot
- Early preparation of funding grants from all co-authors
- Tracking use of research infrastructure, equipment and facilities
- Security control issues

- Institutional grant systems support PIDs
- Research management software to correctly model research

Journal selection stage - sticky notes:

- Global checker tool (using PIDs)
- Funder policy publisher agreements
- Tools to help authors meet grant/funding requirements
- Guidance/clarity of options and implications

Article submission stage - sticky notes:

- Manuscript systems
    - Funding information collected for all authors
    - For each author Grant(s) ID(s), Project ID (RAID), core funding 'ID'
- Pull IDs from Crossref, ROR, ORCID DataCite, RAID
- Each author has an ORCID
- Consistency across submission systems
- Publisher consensus and submission system consensus
- Linked IDs global registry
- New grants DOI registry Crossref
- ROR for organisations and funders
- OA Switchboard
- Plug-ins
- Multiple language support for funder names
- ROR

Link to related entities - sticky notes:

- Datasets
- Equipment
- Preprints
- Protocols
- Data management plans
- Integrity checks
- Including IDs early - e.g. DMPs, preprints, version links
- Pre-registration

Review - sticky notes:

- Before review, all identifiable information is anonymised (optional)
- Advocate for transparent peer review
- Register reviews with Crossref DOI
    - (Editor's note: DataCite can also provide DOIs for reviews)
- Integrity checks

Acceptance - sticky notes:

- Notify all entities
- Notify COAR

- Preprint is noted as accepted, metadata updated
- Pending publication (publisher) DOI?
- Rights retention
- Versioning
- Notify super switchboard

Repository deposit - sticky notes:

- Minimum requirement for AAM deposit in repositories
- Metadata of grant funding information
- Agreed metadata schema for AAMs in repositories
- Repositories needs to accept funder and grant IDs
- Funder notified i.e. DataCite search
- PIDs funder and grant
- Complete metadata
- Automatic deposit to repositories from submission systems
- Feedback loop to ensure complete metadata

Publication - sticky notes:

- VOR to Crossref or update version
- Funder acknowledgements and IDs in VOR
- Notification back to funder and institution
- Retractions or conflicts of interest linked to complete article metadata
- Funding information in VOR and AAM text plus metadata
- Resolvable links to grants and funders
- Validation
- Feedback loop to ensure complete metadata
- Standard structured text approach across funders globally

Reporting and analysis - sticky notes:

- Metadata is already accurate here
- High coverage of systems/support for standard approach (institutions)
- Openly available data/DOIs
- Grants that did not result in publications or outputs?
- Need COUNTER-like definitions of what is counted
- Standard structured approaches across funders
- Benchmarking
- Consistency, deduplication
- Funders can pull the data themselves instead of push from repositories
- Entity links e.g. people, organisations
- Linking multiple versions
- Open metadata
- Interoperability

Next steps

- This meeting formed part of the preparation/update for a second Grant IDs workshop convened on 9 November
- Work for UKRI to do to motivate and support progress
- Other funding bodies are also interested in addressing this problem and there needs to be coordination. Consistency of action would be useful. Crossref to help? Practice is currently inconsistently applied across funding bodies
- Explore opportunities for the Publications Router and other services to have a role

Helpful links and resources

- Publishers and funders convened a workshop in June 2023 to pinpoint challenges in creating funding data: https://www.crossref.org/blog/open-funding-metadata-community-workshop-report/

A6: Summary of UKRI virtual workshops on Grant PIDs

Recommendation

This virtual workshop spanned two sessions and focused on MoreBrains' recommendation for UKRI to work with all relevant stakeholders to raise awareness and understanding of grant DOIs and discuss a pathway to adoption and consistency, to ensure that grant DOIs can form part of the evidence base required for efficiently demonstrating alignment with the policy requirements, and for monitoring and evaluating its impact and effectiveness.

Meeting dates

Session1: 4 October 2023
Session 2: 9 November 2023

Session one

Attendees

- Stacey Burke (SocPC/The Physiological Society)
- Matt Buys (DataCite)
- Matt Cannon (Publshers Association/Taylor & Francis)
- Dominique Capostagno (NIHR)
- Melissa Harrison (JATS/Europe PMC/EMBL-EBI)
- Ginny Hendricks (Crossref)
- Hannah Hope (Wellcome)
- Hylke Koers (STM Association)
- Catriona MacCallum (OASPA/Hindawi/Wiley)
- Ben Ryan (EPSRC)
- Aditya Sehgal (Publishers Association/Elsevier)

Facilitators

- Sara Ball (UKRI)
- Josh Brown (MoreBrains)
- Phill Jones (MoreBrains)
- Fiona Murphy (MoreBrains)

Notes

NB: The views summarised are those of participants and do not necessarily reflect the views, priorities and policies of UKRI.

This presentation was given by Sara Ball, Phill Jones, Josh Brown, Melissa Harrison, Ginny Hendricks, and Matt Buys. The scope of this element of the policy was confirmed to be journal articles, and a series of exercises were conducted on a virtual Miro board. The main areas of discussion were:

- Grant PIDs are at an early stage in their maturity. Sector-wide consensus is required to move things on
- Reporting from researchers is currently late and incomplete. Currently processes are manual, so more automation is required
- Instigating grant IDs would tell a publisher what the funder's OA policy requirements are. It would also flag up when there is a co-funding situation that could require mediation between policies
- It was acknowledged that Grant PIDs and Project PIDs (RAIDs) are closely connected and that enabling systems to provide auto-updates would be very valuable. For clarity, grants and projects are seen as two separate entities, and there was a dedicated workshop on project PIDs on October 13th
- There was a discussion regarding at what point in the process a Grant ID should be applied. There has to be some sort of application number (which ISN'T a PID) that is used to manage all the applications within the grant management software. The parallel is publishers' MSIDs that aren't DOIs. For the latter use case, DOIs are attached at some point between acceptance and publication. It would make sense to do something similar with grants
- It would be a huge benefit if funders started routinely using Grant IDs. There is an assumption among funders that their current numbering systems are unique and discoverable, but PID prefixes and suffixes would clearly link funders to their funded programmes
    - There is a need to ensure that the Grant IDs and ORCIDs in question are related to specific research outputs. There can be multiple awards and outputs that inter-relate, so these distinctions need to be clear
    - There is a need to identify - and highlight - quid pro quo instances so that funders will be incentivised to provide information that will help publishers, and vice versa
- One use case is the funder pushing the award information to the research organisation at the point of award. This would include personnel details (researcher, technicians and so forth) and their ORCIDs (including technicians etc) so if they're involved in the future, reviewers only need to check ORCIDs.

- Via the Freya Project and working with Crossref, Wellcome, and PLOS, Europe PMC (as part of EMBL-EBI) has built an infrastructure to register DOIs for grants awarded by those funders which support Europe PMC, via Crossref
    - As part of this project, Wellcome joined Crossref as a member, and Europe PMC performs the technical work to register grant DOIs with Crossref for Wellcome. This workflow is extensible to other Europe PMC funders
    - Grant DOIs are registered using an automated pipeline, but there is a submission form provided by Crossref for users if required
    - Each grant needs a url for a landing page, and this link must be submitted as part of the grant DOI registration metadata. Europe PMC hosts this page for Wellcome
    - There also needs to be a funder ID as well as other optional and required metadata
    - PLOS coordinated with Wellcome-funded authors to populate their articles with Grant DOIs
    - If the funder (Wellcome) provides the PI's ORCID with the grant information, Europe PMC will include this in the Grant DOI metadata
- ROR acts as a curated registry, so anyone can make a suggestion to add a new record to the registry or correct an existing one and it will go through to a curation committee
- DataCite has launched a Grant ID pilot that leverages Crossref's schema of Grant IDs. This means that metadata can be registered within DataCite's standard tools. There are plans to embed this as part of the existing fee model with DataCite's members. (See this link for information on how a data management plan can be used with this functionality.)
    - This will enable enquiries to move either upstream or downstream of their original entry into the network of PIDs

Technical challenges:

- The schemas need further development
- There are concerns with Grant PID, how much data cleaning needed to make these powerful? Don't understand institutional systems, so not sure what we'd be relying on?
- If grant IDs could be integrated with ORCIDs to flow into a profile, that would fit very well with ORCIDs 'trust markers' concept

Incentive challenges:

- Potential benefits include metadata re-use, automation of processes, strategic analysis, a genuine source of truth
- There's potentially a need for use cases from the perspective of each stakeholder within the system to demonstrate the net payback element of registering PIDs within their own remit
    - Regarding grant IDs, there is an incentive for publishers to do something around it if they have access to successful grant applications that will eventually be associated with papers

What UKRI can do to help

- UKRI could join Crossref and start issuing Grant IDs. As well as providing the information required to populate the other systems, this would greatly accelerate the growth of trust in the system. It would indicate the direction of travel and give other stakeholders the confidence to invest

- UKRI can continue to convene stakeholders to motivate and support progress

Session two

Attendees

- Matt Buys (DataCite)
- Matthew Cannon (Publishers Association/Taylor & Francis)
- Dominique Capostagno (NIHR)
- Melissa Harrison (JATS4R/EMBL/EBI)
- Catriona MacCallum (OASPA/Wiley/Hindawi)
- Helen Nolan (NIHR)
- Ed Pentz (Crossref) - left at 60 minutes
- Aditya Sehgal (Publishers Association/Elsevier)

Facilitators

- Josh Brown (MoreBrains)
- Rachel Bruce (UKRI) - joined at 60 minutes
- Phill Jones (MoreBrains)
- Fiona Murphy (MoreBrains)

Notes

UKRI and MoreBrains shared a presentation that outlined the background for the open access policy, previous discussion of grant PIDs, and an update on the funding acknowledgement meeting on 2 November.

The group then moved on to a two-part exercise using a shared Miro Board.

Group exercise one:

The first exercise identified actions that need to be taken by various actors at each stage of the research lifecycle to deliver the benefits of grant PIDs.

Stage 1: grant application and review

Researchers need to:

- Register ORCID IDs, share these with funders and grant read/write permissions to the funder grant management system
- Share details (ideally with PIDs associated via automated tools) for other researchers involved, affiliations, previous/related grants and projects, Data Management Plans etc.

Funders need to:

- Join Crossref or other grant PID provider, and register PIDs for grants upon award
- Leverage PID registries (e.g. ORCID) to pre-populate key information wherever possible (e.g. affiliation switch associated PIDs such as ROR)
- Include all collected PIDs in grant PID metadata and ensure metadata is complete and remains up to date
- Use grant and funder PIDs in all correspondence relating to the grant

- Issue clear guidance and requirements on the use of grant PIDs in funding acknowledgements

Institutions need to:
- Support researchers in populating their ORCID records and educate them in why/when to grant permissions
- Use grant and associated PIDs in internal systems (e.g. CRISs, repositories)

Stage 2: research in progress

Researchers need to:
- Include grant PID with any research output
- Use Grant PID in all correspondence with institution and/or funder

Funders need to:
- Use Grant PID in all correspondence with researcher and/or institution
- Proactively search for research outputs based on grant PIDs to prepopulate any annual progress reporting
- Keep grant metadata updated about people information using DOI and ORCID
- Create public landing page to showcase ongoing outputs associated with a grant

Institutions need to:
- Update funder if people are being added or removed from funded projects (with ORCID)
- Include grant PID with any research output
- Support interactions with external infrastructures (disciplinary etc.)
- Connect resources and outputs to grants and make it bidirectional

Other actions needed:
- Relevant repositories (e.g. Zenodo, preprint servers) should provide tools to link grant PIDs to outputs/work

Stage 3: publication

Researchers need to:
- Provide grant PID in all publications

Funders need to:
- Encourage/educate authors to include Grant PID in publications

Publishers need to:
- Facilitate ORCID log in, and authors be presented with linked grants to prepopulate other metadata/information fields
- Include grant PID in publication metadata and provide human and machine-readable funder acknowledgement section in all papers that includes grant PIDs
- Validate grant ID at submission
- Deliver Grant PID linking information when registering article DOI

Repositories need to:

- Include grant PID in publication metadata
- Enable programmatic access to pull grant PIDs automatically or via lookups or search/link tools
- Cross repository notification system to expose different versions of the same thing linked via person AND grant PIDs

Institutions need to:

- Encourage/educate authors to add grant PIDs to publications

Stage 4: reporting

Publishers and repositories need to:

- Automatically share usage data for other stakeholders (downloads, citations, altmetrics)

Funders need to:

- Collect Grant PID usage from DOI registries
- Share instances where outputs were sourced using Grant PIDs and not annual reporting mechanisms
- Proactively communicate/promote research outputs
- Chase up researchers who have not complied

Institutions need to:

- Collect grant PID usage from DOI registries
- Associate institutional reporting requirements with grant PID
- Repositories to more fully adopt standards around metadata and endpoints

Group exercise two:

The second exercise explored key dependencies for the actions identified in exercise one to be delivered consistently.

A critical dependency for grant PIDs to be used to support UKRI's open access policy would be for UKRI to start to register grant PIDs. This was used as an example to help to structure the exercise.

What needs to happen for…

1) Notifications of publication events to be shared widely?

- There needs to be a standard for notifications and an agreed list of events that merit a notification being sent to ensure the integrity and accuracy of data
- Need to ensure that notifications get sent to funders and institutions
- Organise grant PID metadata collection from outputs works via those systems
- Provide easy access/connectivity to grant PIDs to ensure grantees use the grant PIDs in their outputs
- Integrity checks to pick up fraud/duplications (such as identity theft)
- Needs to work for non UKRI authors/coauthors

- Secure way to subscribe to notifications and adhering to other standards such as GDPR with rules about who can update each event and who can see/access each notification
- System requires transparent governance to support trust and adoption across competing organisations

2) UKRI grant PIDs to be updated systematically?

- Have a mechanism for institutions/researchers to update UKRI and for UKRI to update grant metadata
- Explore connectivity between grant management systems linking to Grant PID host
- Clarity on co-funded organisations' responsibilities
- Clear policy on what needs to be updated
    - New hires?
    - Changing partners?
    - Changes of grant (e.g. change in value?)
- Input from those managing grants at the moment
- Explore connectivity between grant management systems linking to Grant PID registry

A7: Summary of UKRI Virtual Round Table on ISSNs

Recommendation

The focus of this virtual round table was MoreBrains' recommendation that UKRI: encourages the repository community to work on mechanisms and workflows to improve levels of ISSN coverage; promotes the use of RIOXX or the DataCite schema; provides specific instructions on what metadata fields need to be included and how they should be formatted; and works with Crossref and Publications Router to develop ways to supplement repository metadata.

Meeting date

21 September 2023

Attendees

- Steve Byford (Jisc/Publications Router)
- Petr Knoth (CORE)
- George Macgregor (UKCoRR/University of Glasgow)
- Valerie McCutcheon (ARMA/University of Glasgow)
- Patricia Feeney (Crossref)

Facilitators

- Sara Ball (UKRI)
- Phill Jones (MoreBrains)
- Fiona Murphy (MoreBrains)

Notes

NB: The views summarised are those of participants and do not necessarily reflect the views, priorities and policies of UKRI.

Following this presentation by Sara Ball and Phill Jones and after confirming that the scope of this element of the policy is journal articles, the main areas of discussion were:

- The need to include ISSNs is mandatory for both publishers and repositories. The decision was taken to promote the use of schemas/application profiles so UKRI can have access in a predictable way. However, there are still a number of articles which don't have ISSNs associated with them, including:
    - A few journals don't have ISSNs
    - According to Crossref, there are some other gaps such as Japanese articles and law journals
    - Conference proceedings, preprints, slides, theses
- If we accept that there are some edge cases, the key questions become: how to get more ISSNs into repositories, and then how to get more ISSNs out of the repositories (and have fewer gaps)? (CORE can read from dc:source.)
- The ISSN is a critical identifier so you would expect institutions to record it (But when in the workflow?)
- There are resourcing issues for institutional repositories. Much of this data is currently manually added, so dependent on having staff available to do the work.
- When publications are first put into the repository, researchers don't have an ISSN. ISSNs must therefore be added later. Repositories with fewer resources are unable to monitor changes in ISSN status and can't update records
- Researchers don't necessarily recognise the importance of the ISSN. It's likely that it will be more useful to work with institutions rather than researchers - systems and repositories should support ISSNs as standard
- There are always limits in the way data are exposed and captured. Some repositories use RIOXX, others rely on Dublin Core with patchy compliance with OpenAIRE guidelines. ISSN can be added into any of these
- Mechanisms for retrieving ISSNs for repository records are variable. Repositories use a variety of datasets including Scopus, publication router etc, leading to patchy, and sometimes conflicting data

Current capabilities

- Articles passing through the Publications Router should always have an ISSN
- For articles NOT passing through the Router, Eprints did have a plug-in which enabled the user with the DOI to pull down additional information from XR. This could be (re)commissioned to support the policy. (https://eprints.gla.ac.uk/303064/: example from router with ISSN)
- If the DOI to a VoR is given, it should be possible to retrieve the ISSN automatically from Crossref or another metadata registry
- ISSNs can be pulled across from ORCID records

- EuropePMC might be a special case where meaningful compliance reports could be extracted. It has higher levels of functionality than most institutional repositories

Possible solutions:

At the point of deposit for publications deposited after VoR publication

- CORE is considering a tool/API that would take the article title/DOI and then prefill fields for validation at the point of submission to the repository
- An auto complete system would involve entering the DOI and pinging the Crossref API, which would then pull the ISSN into place. Some institutions are already using this method

To update existing records, and those where the manuscript is put in the repository before it is published and metadata like ISSN is not known

- Publications router can be used to push data to repositories
- It is technically straightforward to retrieve the data on a schedule eg once a week via Crossref's API and update the records. A plug-in could be developed to support repositories in this
- CORE could provide a service to aggregate ISSNs. This could take the form of
    - (a) a post-deposit service that finds ISSNs, makes them available in the CORE Enrichments module from which they get propagated to repositories and/or
    - (b) during deposit, i.e. through the provision of a service that recommends and prepopulates the ISSN field for the depositor based on the already entered metadata

Items without ISSNs

UKRI needs to be able to treat different types of publications differently and ascertain what standards to hold them to.

Repository and research managers' problems, and what UKRI could do to help

- Universities need to receive the relevant metadata earlier than when the information reaches CORE or PubRouter. When the publishers are certain, for instance? Currently institutions are collecting information directly via publishers' dashboards, and it is still error prone. UKRI could facilitate the conversation about how to connect publisher systems to repositories earlier in the process
- Advocate for automated tools. Emphasise the importance of recording what we've been discussing today. The people administering the systems and the institutional owners may not be aware of its importance
- Focus on what can be done to make the process(es) more efficient rather than just paying people more to do things
- A simple to follow toolkit for institutions to recommend places to go for the information. We shouldn't push the idea that more staff are necessarily needed, but could make things simpler with these sorts of resources
- Develop a toolkit to support institutions to work through options for populating repository metadata?

- UKRI could fund the creation of plug-ins for Eprints and patches for DSpace. The proprietary systems are a challenge but may have the capacity to invest themselves
- Explain to the publishers what the value and reach of their metadata is, and involve them in the conversations

Outstanding questions

- Is there a need to articulate this more directly to publishers? And for the message to come via the funders?
    - The publishers weren't directly involved in this conversation but were present by proxy via Crossref.

Barriers

- Some institutional repository owners work in research management rather than repository management functions and aren't aware of the importance of these metadata and workflows. Education is therefore required – and a toolkit will also help.
- The REF. The priority for the institutions is already shifting to the next REF Immediately after a REF cycle is the best time to engage, with the next opportunity being 2029
- Is it possible to frame this as a positive for managing REF? A lot of what we are discussing should help information move better through the system, which in turn should help save time and resources for the institutions that are working on REF
- Some uncertainty as to which repositories fall into this space - do the UKRI specialist repositories, such as NERC, qualify? (We think they do)
- Suppliers are not incentivised to participate in solutions to this problem. But if the solution could be framed to meet both publisher and institutional goals then it could be a win-win
- There will be edge cases, for example journals that don't issue DOIs. Manual back-up processes will always be needed for minority cases

A8: Summary of UKRI In-person Workshop on Licensing

Recommendation

This half-day in-person workshop focused on MoreBrains' recommendation that UKRI work with the community and relevant stakeholders and infrastructure providers to encourage the adoption and use of the NISO ALI metadata scheme for licence reporting across the publishing and repository sectors. In collaboration with the community, MoreBrains recommended the development of advice and best practice documentation detailing which metadata fields are required for compliance and how to adopt the schema.

Meeting date

2 November 2023

Attendees

- Todd Carpenter (NISO)
- Eleanor Dumbill (CoSector)

- Ginny Hendricks (Crossref)
- Kelly Hetherington (UKCoRR/Durham University)
- Leslie Lansman (ALPSP/Springer Nature)
- Catriona Maccallum (OASPA/Wiley)
- Agustina Martinez-Garcia (UKCoRR/Cambridge University)
- Leila Moore (Publishers Association/Wiley)
- Iratxe Puebla (DataCite)
- Sandra Townsend (eJournal Press/Wiley)
- Nathan Westgarth (Aries/Elsevier)

Facilitators

- Sara Ball (UKRI)
- Josh Brown (MoreBrains)
- Phill Jones (MoreBrains)
- Fiona Murphy (MoreBrains)

Notes

NB: The views summarised are those of participants and do not necessarily reflect the views, priorities and policies of UKRI.

Following this presentation by Sara Ball (UKRI), Fiona Murphy (MoreBrains), Todd Carpenter (NISO), Leila Moore (PA/Wiley) and Sandra Townsend (eJP/Wiley), and after confirming that the scope of this element of the policy is journal articles, the main areas of discussion were:

- Some schemas don't support all the metadata fields
- Does UKRI need to implement a timeline for getting schemas extended, and then for other stakeholders to roll out the revised schemas?
- What are the timelines, barriers and what support or guidance is needed? It has to be possible for people to comply and possible for them to show how it's done
- What existing good practices are already there and being used?
- The notion of author choice around licensing is hugely important. There are some psychological factors (a recent survey showed that if they are offered a more restrictive option first, authors typically will NOT choose CC-BY of their own accord)
- There is a lack of understanding of what the licences are
- There is a lot of work going on to update systems in response to a variety of demands and requirements. Considerations around this policy could be too much for tiny publishers
- Sometimes publisher and funder requirements don't match
- There is a need to move to clear expression of licences and pinpoint exactly what needs to be supplied to authors

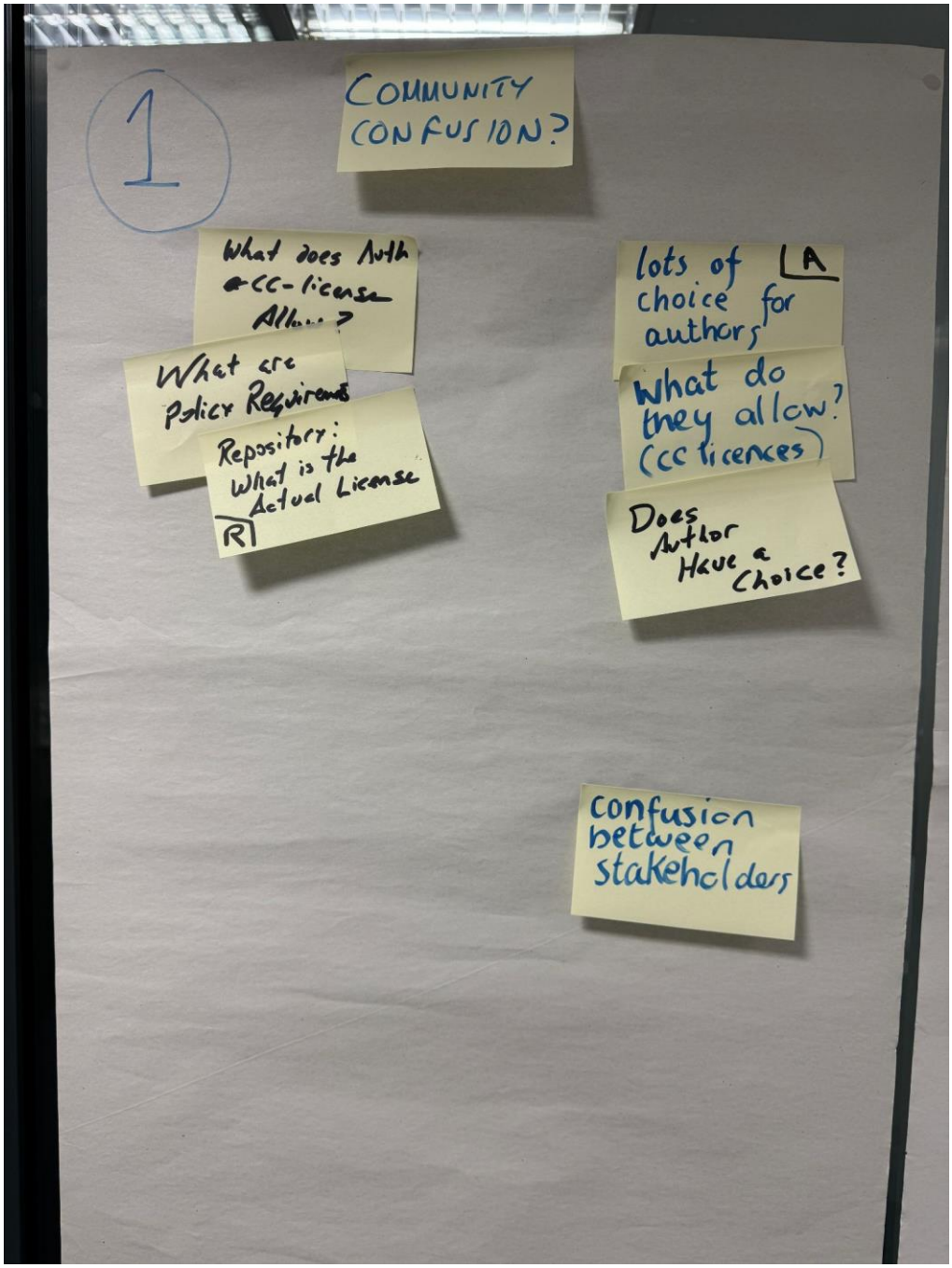
A three-part group exercise was then conducted. Delegates divided into two groups which went through the same set of questions independent of each other. In the first step we gathered issues with efficiently collecting licensing information. In the second, we collected suggestions for what can be done to make this process both take place and work well. The final step - conducted in

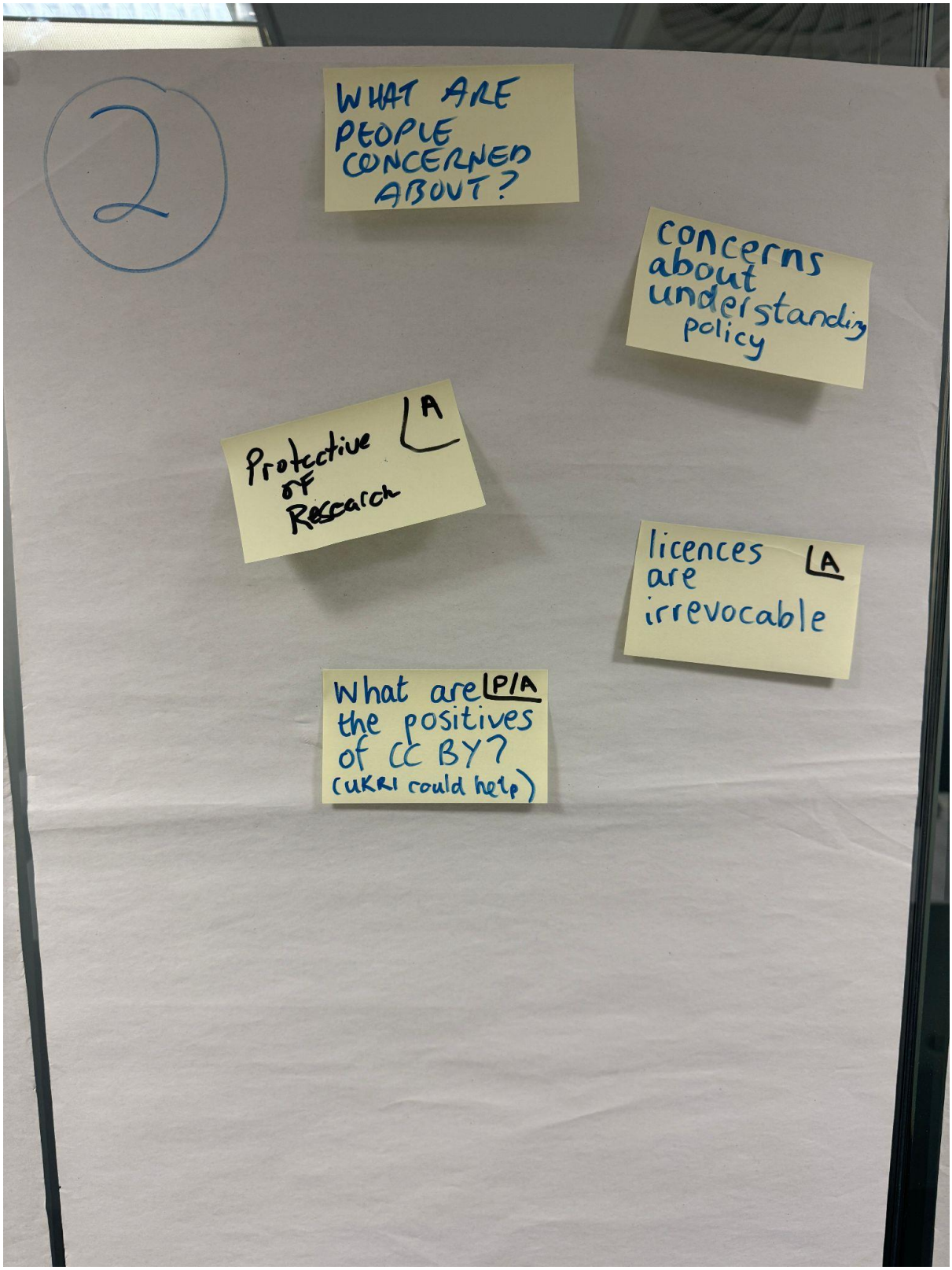
plenary - consisted of evaluating the suggestions according to perceived effort (rated from easy to hard) versus impact (high to low). The final outcomes are listed below:

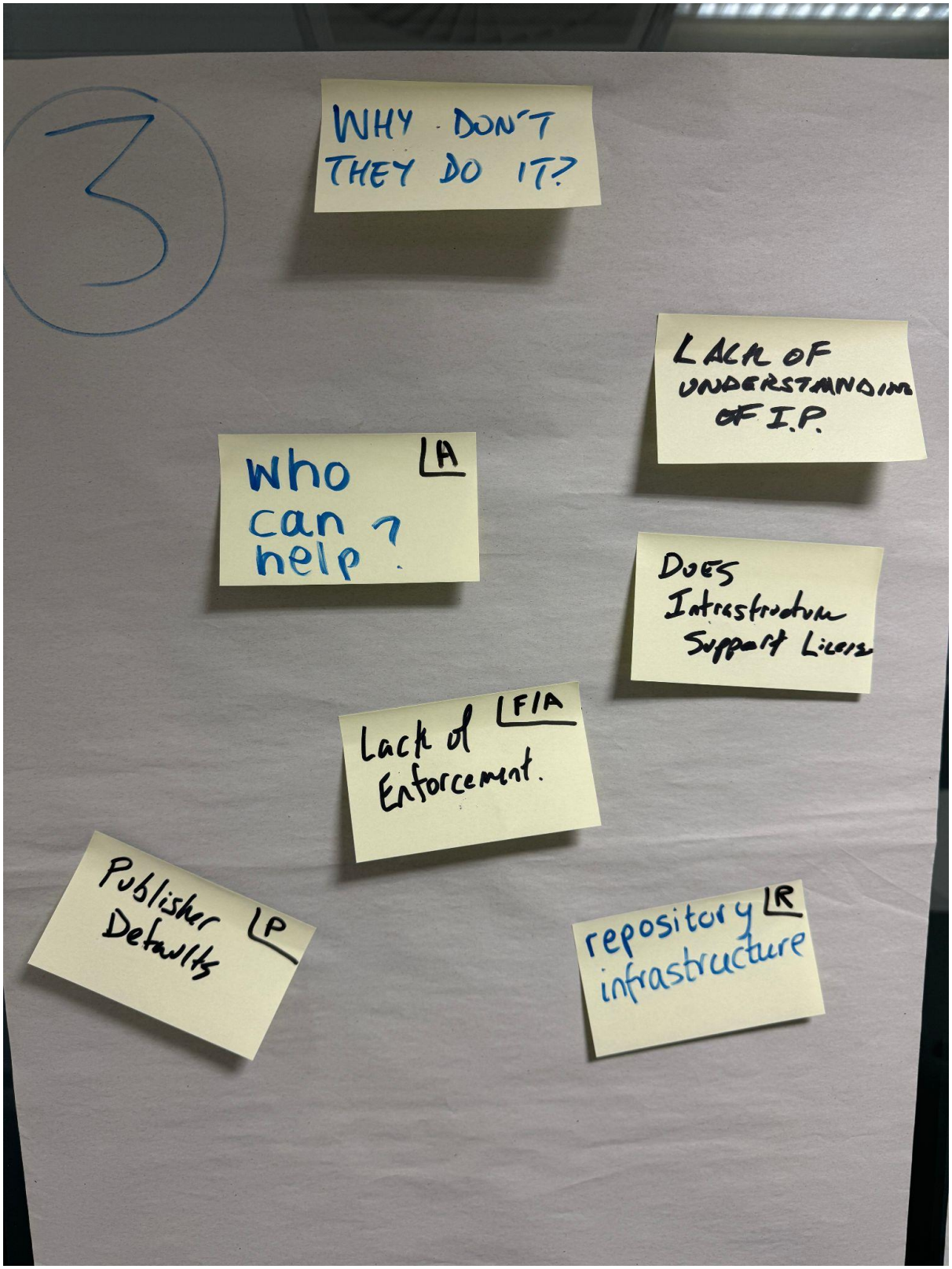
- Easy and High Impact:
    - Funders give clear recommendations to repositories
    - Require that licences are machine readable
    - Advertising campaigns for: CC-BY, how researchers will benefit from the policy, awareness of better outcomes, publisher awareness
    - Provide a clear message that new technical approaches are not required as ALI exists
- Hard and High Impact
    - Machine readable funder policies
    - Limit/restrict the number of bespoke licences used by publishers
    - Enforcement of the policy by UKRI
    - Policy alignment between funders and institutions
    - Publishers provide more guidance and more consistent guidance to researchers
    - Clarify the different CC licences
- Easy and Low Impact
    - Clarify copyright and copyright transfer
- Hard and Low Impact
    - Refine licence schema
    - Integrate into submission systems

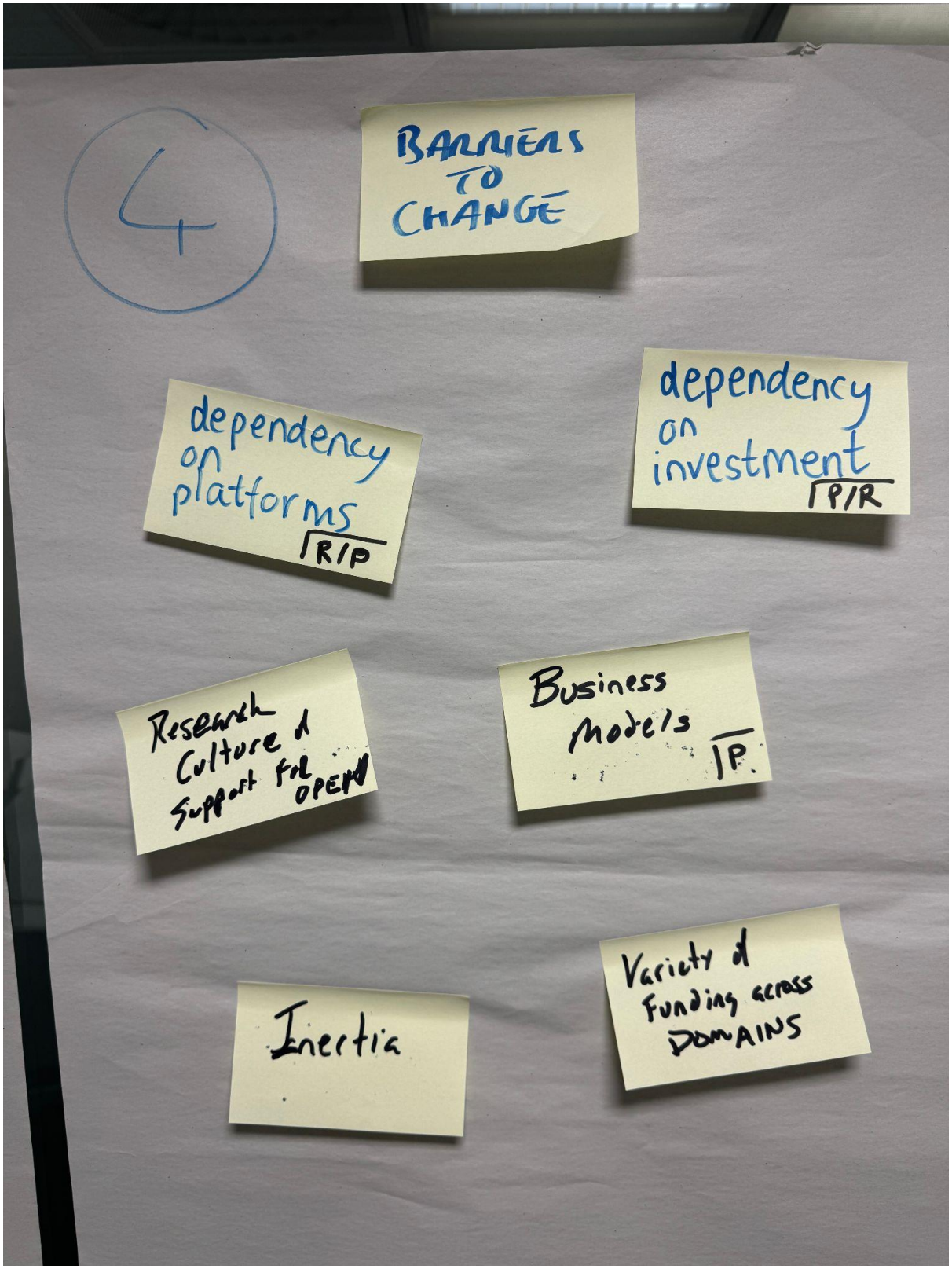
The groups did not achieve consensus on the suggestion to ‘develop a massive journal checker tool’. While both groups identified this potential initiative and recognised that it would be ‘hard’ to implement, one assessed its impact as ‘high’ while the other came to the conclusion that it would be ‘low’.

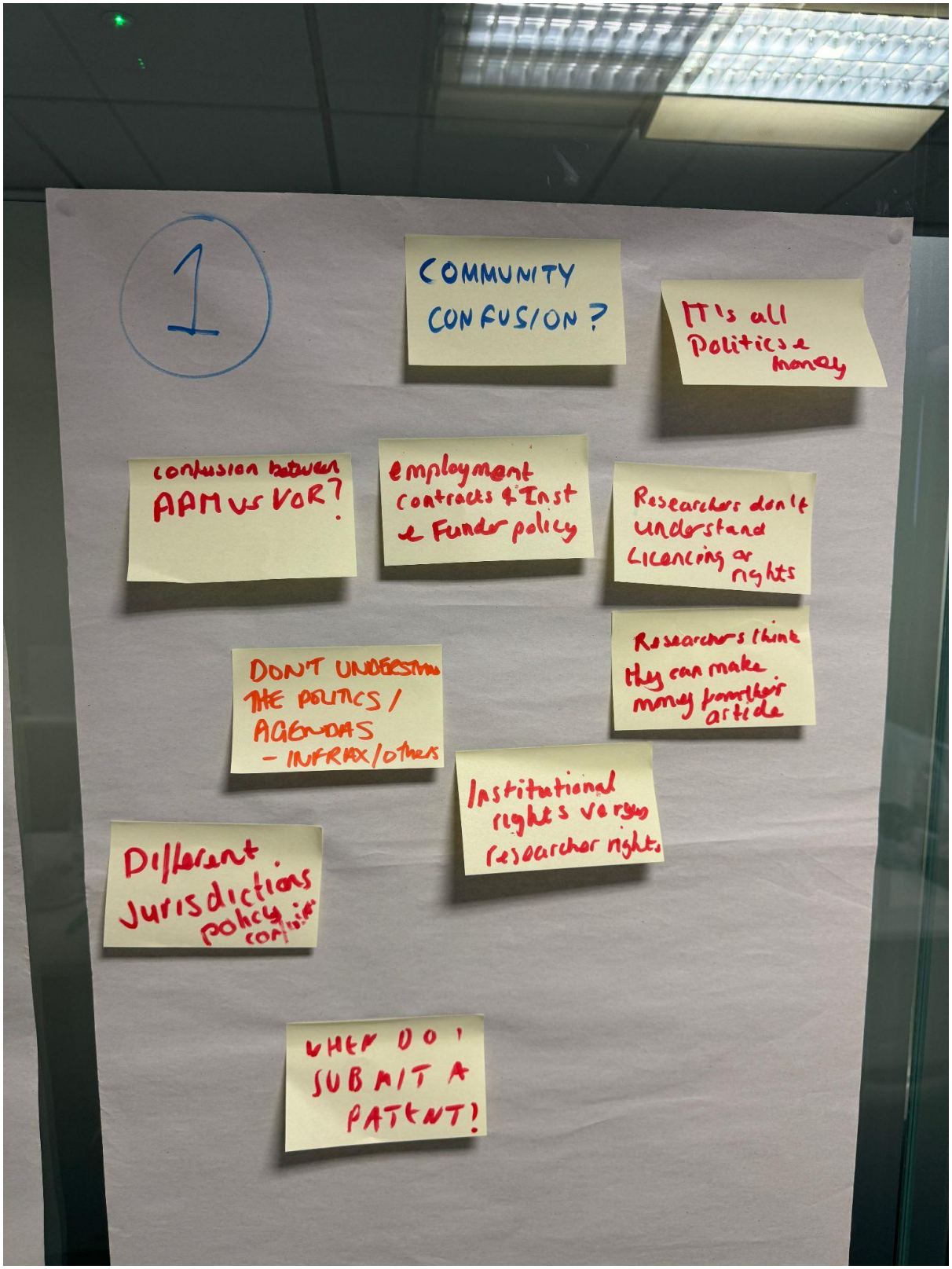
Photographs of outputs from group exercises

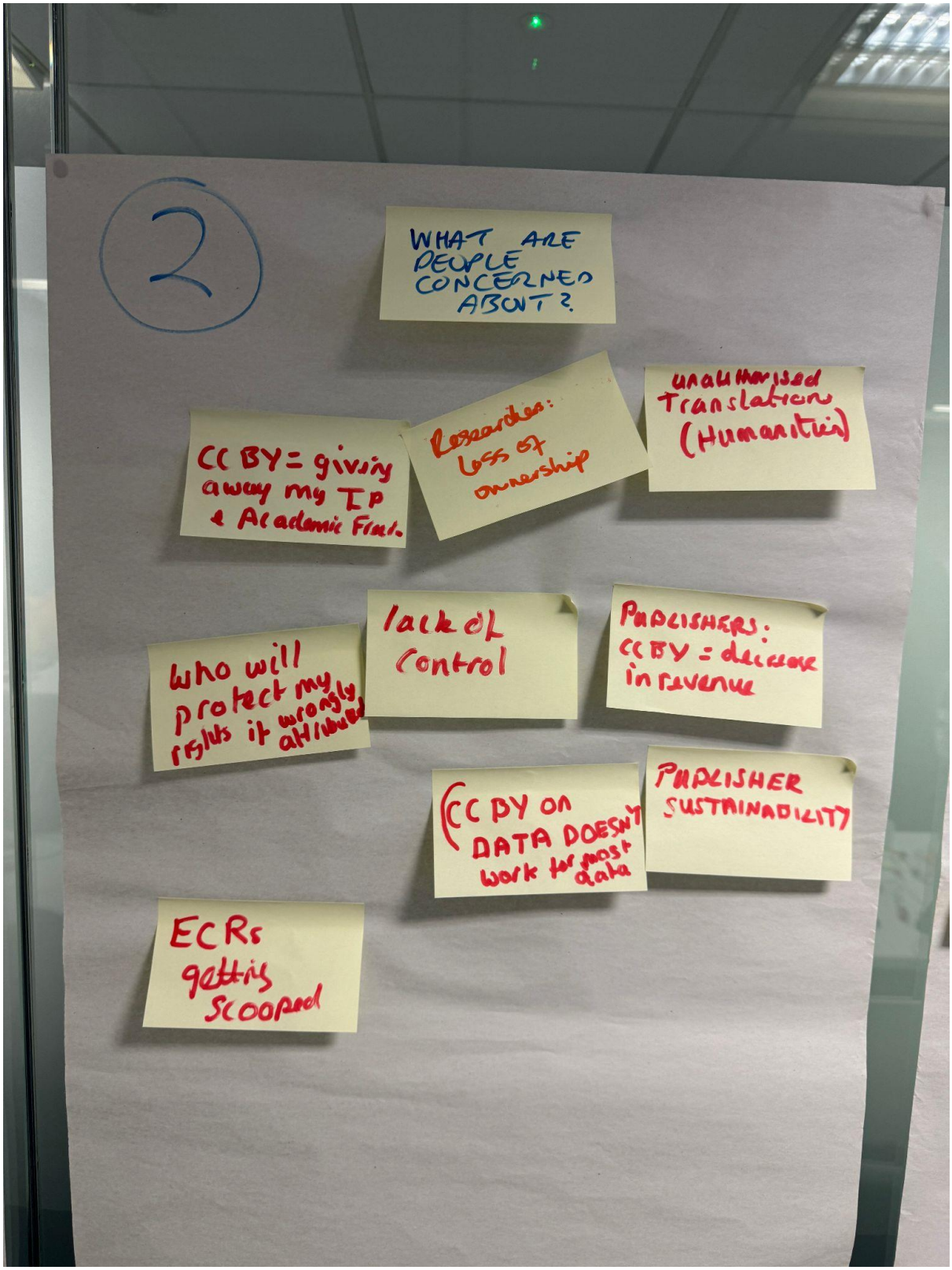


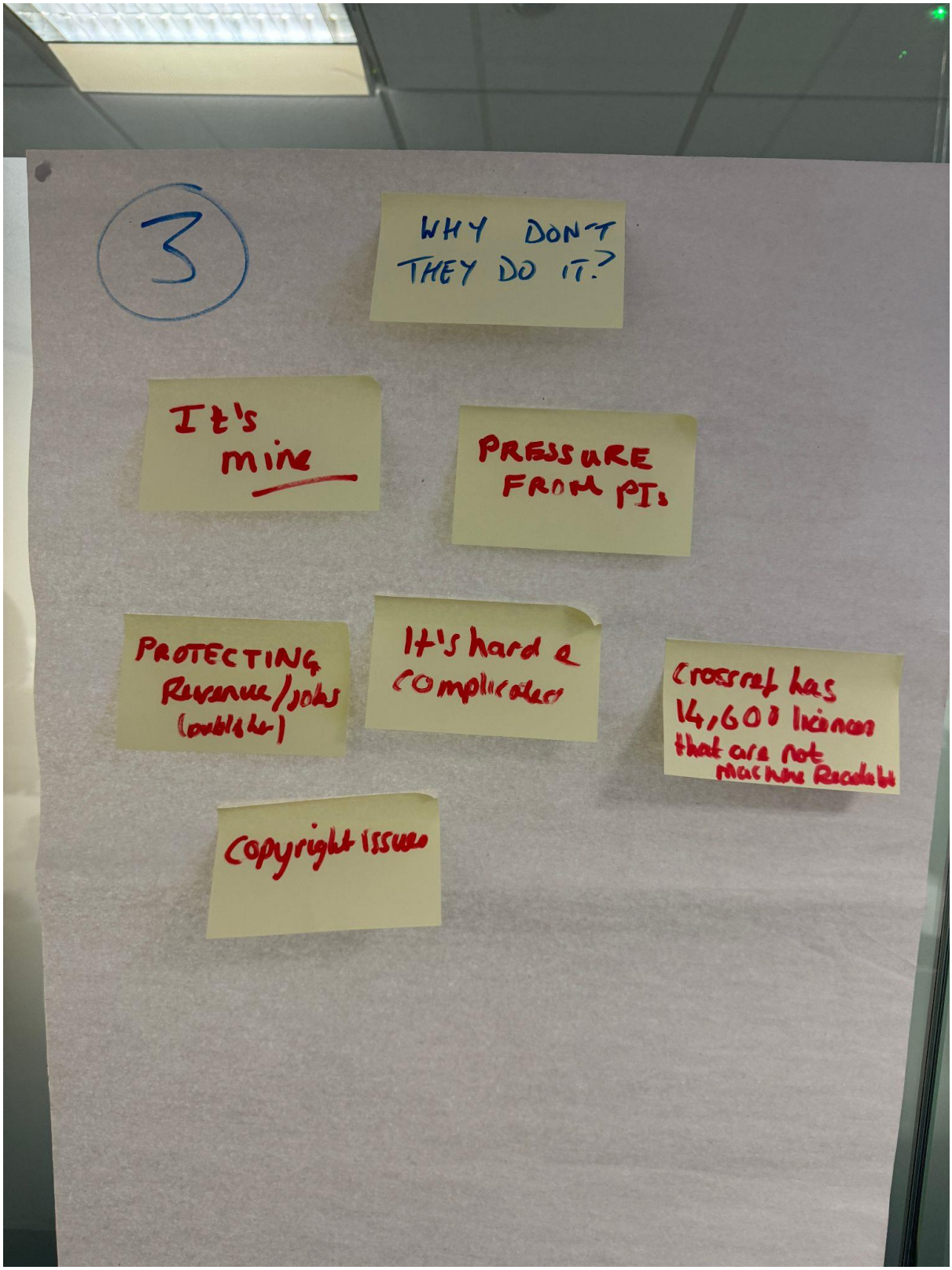

2
WHAT ARE PEOPLE CONCERNED ABOUT?
concerns about understanding policy
Protective of Research
LA
licences are irrevocable
LA
What are the positives of CC BY? (UKRI could help)
LP/A

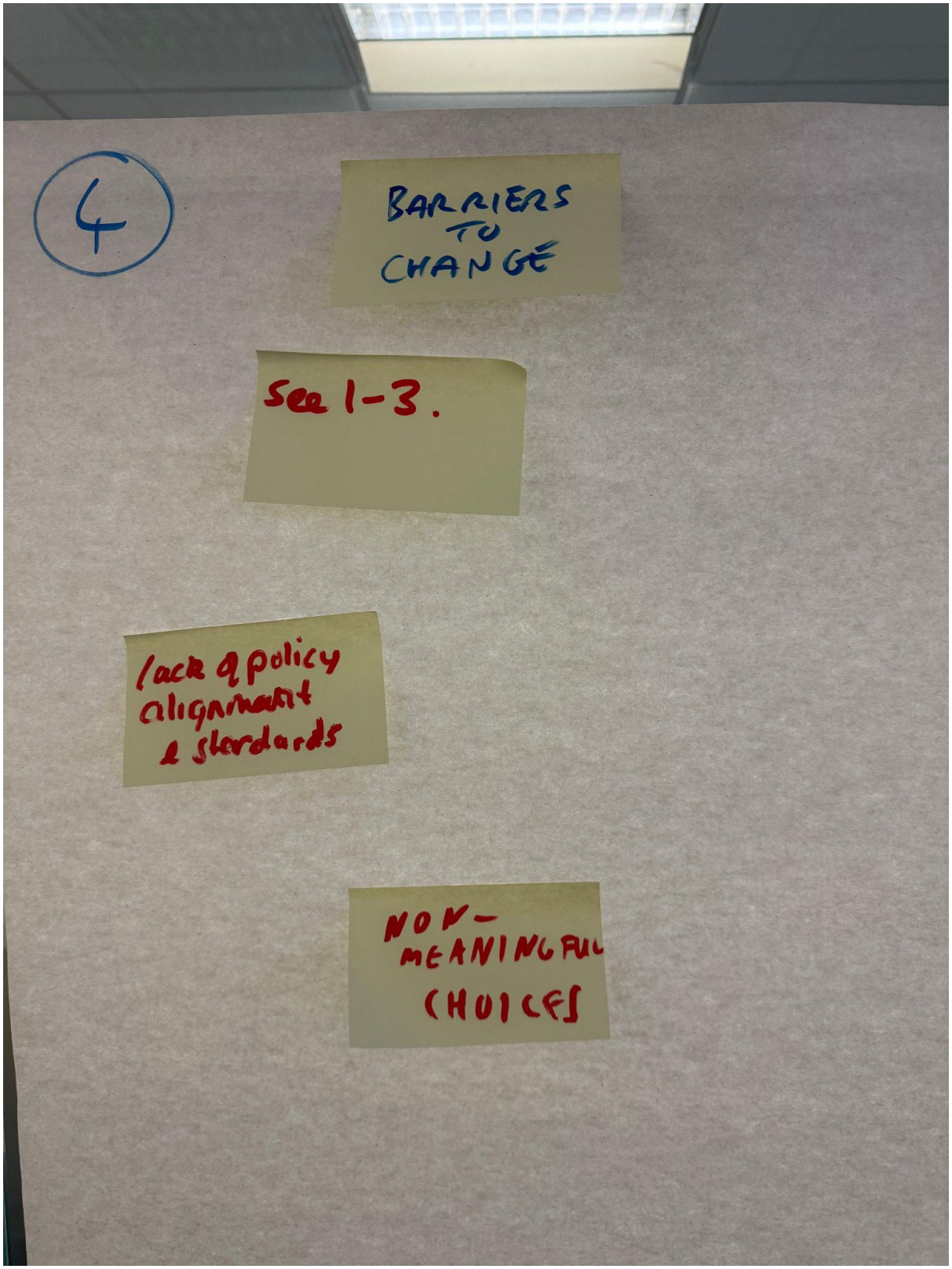

3
WHY DON'T THEY DO IT?
LACK OF UNDERSTANDING OF I.P.
Who can help? LA
DOES Infrastructure Support Licen
Lack of Enforcement. LFIA
Publisher Defaults LP
repository infrastructure LR


4
BARRIERS TO CHANGE
dependency on platforms
R/P
dependency on investment
P/R
Research Culture & Support for OPEN
Business models
P
Inertia
Variety of Funding across DOMAINS


1
COMMUNITY CONFUSION?
IT's all Politics & Money
confusion between AAM vs VOR?
employment contracts & Inst & Funder policy
Researchers don't understand Licencing or rights
Researchers think they can make money from their article
DON'T UNDERSTAND THE POLITICS / AGENDAS - INFRAX/others
Institutional rights versus researcher rights
Different Jurisdictions policy
WHEN DO I SUBMIT A PATENT!


2
WHAT ARE PEOPLE CONCERNED ABOUT?
CC BY = giving away my IP & Academic Freed.
Researchers: Loss of ownership
unauthorised Translations (Humanities)
Who will protect my rights if wrongly attributed
lack of control
PUBLISHERS: CC BY = decrease in revenue
CC BY on DATA DOESN'T work for most data
PUBLISHER SUSTAINABILITY
ECRs getting scooped


3
WHY DON'T THEY DO IT?
It's mine
PRESSURE FROM PI's
PROTECTING Revenue/jobs (publishers)
It's hard & complicated
Crossref has 14,608 licences that are not machine Readable
Copyright issues


4
BARRIERS TO CHANGE
See 1-3.
lack of policy alignment & stardards
NON-
MEANINGFUL
CHOICES

A9: Summary of UKRI Virtual Meeting on OpenDOAR registration of repositories

Recommendation

This virtual meeting focused on MoreBrains' recommendation that UKRI work with the community to develop reporting mechanisms and workflows for cross-checking OpenDOAR registration compliance.

Meeting date

September 20, 2023

Attendees

- Alina Chowdhury (Jisc/OpenDOAR)
- Melanie Heeley (Jisc/SHERPA, Romeo, Juliet/OpenDOAR)
- Azhar Hussain (Jisc/OpenDOAR)
- Petr Knoth (CORE)
- George Macgregor (UKCoRR/University of Glasgow)
- Kathleen Shearer (Confederation of OA Repositories)

Facilitators

- Sara Ball (UKRI)
- Phill Jones (MoreBrains)
- Fiona Murphy (MoreBrains)

Notes

NB: The views summarised are those of participants and do not necessarily reflect the views, priorities and policies of UKRI.

Following this presentation by Sara Ball and Phill Jones and after confirming that the scope of this element of the policy is journal articles, the main areas of discussion were:

- For this and several other requirements, it is important to make sure that metadata can be harvested from repositories through some kind of operational, interoperable and documented mechanism, for example, an API
- There was a call for clarity around why UKRI needs repositories to be registered in OpenDOAR, beyond it being considered best practice. Bearing in mind that currently OpenDOAR doesn't do quality checks and can't be considered as a certification. There may also be repositories internationally that the community considers legitimate, but are not in OpenDOAR
- In its current role as a registry of OA repositories, OpenDOAR does conduct checks and holds certain metadata on the repositories it registers but it does not provide accreditation services. The policy does not specify accreditation, merely registration
- For information to easily flow from repositories to UKRI, or other interested parties, functional API endpoints that work in standardised and documented ways are required. This

requirement is not currently part of the policy, and a future review should therefore specify the technical mechanisms (e.g. standard protocols) for exposing data in the future

- Registries in general could be used as an aggregator of metadata, but in its current form OpenDOAR doesn't do this. There's a need for further work to enable aggregation of metadata. This could technically be done either at OpenDOAR or by a third party, such as CORE. These possibilities require further investigation and collaboration with relevant stakeholders
- UKRI could look at the Plan S requirements with regards to appropriate repositories, which has 6 stipulations, including OpenDOAR registration
- OpenDOAR already provides a health check for repositories with respect to their compliance for Plan S, so it could be enhanced or repositioned to perform an accreditation role for UKRI's policy. OpenDOAR is looking at assigning IDs to repositories that are associated with greater levels of quality checking in the future
- There is a need to have either monitoring or certification of technical capabilities of repositories, particularly around metadata availability and interoperability beyond those that are currently in the policy. Who is best placed to accredit repositories is an open question. UK repositories need to be actively engaged with to ensure they are registered, although those based outside the UK will be more difficult to reach
- It is important to make sure that domains and endpoints as documented in OpenDOAR are up to date so that repository registration can be verified. In many cases, they are not
    a. There is a need for some outreach/comms from UKRI. Many institutions in the UK have retired their repository and replaced it with a CRIS. Those systems are often managed by a different part of the institution (the research office) where there is less familiarity with the need for repository interoperability so have not been part of this conversation to date, and they should be included
- Potential ways to improve institutional requirements might include:
    a. Stronger incentives or enforcement of mandates to keep OpenDOAR up to date
    b. A subset of OpenDOAR for UK repositories might increase community buy-in by creating a national view. COAR's international repository directory works with national and regional organisations that have an interest in keeping repository records up to date
    c. Aggregation systems are relying on the registries and synchronising with them. CORE could query OpenDOAR and automatically input new repositories. This functionality isn't fully reliable yet, but could be strengthened if the data were to be maintained sufficiently to be usable by aggregators
- OpenDOAR recently finished manually reviewing all the UK repositories and welcomes opportunities to collaborate. Whatever UKRI requirements are, OpenDOAR can work to supply the information, but it will need clarity and time to do so
- OpenDOAR currently uses its own identifiers but is interested in moving to formal PIDs. This would need to be investigated. Currently there is no consistency in how repositories are identified
- It was said that you can use either OAI IDs or repository URLs to identify whether a repository is in OpenDOAR. With the caveat that the domain for the repository must be up to date, if you use the URL, likewise, the OAI ID must be supplied if that's what's used for verification

- DataCite doesn't specify file level data, just the PID. Any file content hangs off that so there are cases in which a URL for a repository landing page may be provided, but the specific AAM cannot be verified. Further technical consultation is required to understand the specific workflows that could be supported
- Options for building community engagement include:
    a. UKCoRR membership and ARMA members are at the front line so need to be involved. CRIS user groups are probably the weak spot
    b. It's important that UKRI helps to lead the engagement as the other groups pay heed to them, they're in a position to help with coordination, and there's a case for building internal engagement within UKRI among senior people
    c. Registration in OpenDOAR has been voluntary to date, with no real push to join. If there is a clear use case from UKRI that will change
    d. A roadshow, and more outreach generally will get more engagement
    e. Keep answering the 'why' questions and providing clear use cases
- CORE could produce a spreadsheet of the repositories registered in OpenDOAR. During an audit it might be advisable to check for articles NOT deposited in the UK repository system. But 90+% would be covered
- The model could be flipped so if a repository isn't registered, that could trigger a request for a report and then the registration could be organised
- A virtuous circle could be created where stakeholders engage with the community, and the mandate is instilled, then better data emerges about the gaps and the missing repositories can be invited to register

Technical points towards working solution(s)

- UKRI can ping the OpenDOAR API, and it will say yes or no, simply
    - Technical investigation is required to understand the best way to identify the repository. This could be through OAI ID or URL. In both cases, data availability and accuracy need to be addressed

Outstanding questions

- How is the specific motivation for this requirement best articulated in order to encourage community buy-in?
- What is the best workflow in order for UKRI to assess if an AAM is present in a repository?
- What is the best workflow to test if a repository is registered in OpenDOAR?

A10: Summary of UKRI In-person Workshop on ORCID

Recommendation

This full-day in-person workshop focused on MoreBrains' recommendation that a phased approach to implementing the requirement for ORCID IDs for UKRI funded authors should be developed in consultation with key stakeholders.

Meeting date

November 3, 2023

Attendees

- Christopher Brown (Jisc)
- Tom Demeranville (ORCID)
- Will Fyson (CoSector)
- Melissa Harrison (EMBL-EBI)
- Rory McNicholl (CoSector)
- Catriona MacCallum (OASPA/Wiley/Hindawi)
- Sam Perkins (Sage)
- Philip Reimann (Clarivate)
- Aditya Sehgal (Publishers Association/Elsevier)
- Nathan Westgarth (Aries/Elsevier)

Facilitators

- Sara Ball (UKRI)
- Josh Brown (MoreBrains)
- Phill Jones (MoreBrains)
- Fiona Murphy (MoreBrains)

Notes

NB: The views summarised are those of participants and do not necessarily reflect the views, priorities and policies of UKRI.

The workshop opened with presentations by Sara Ball on the UKRI Open Access policy and the rationale for the requirement for ORCID IDs for UKRI-funded authors; by MoreBrains on sector readiness to collect ORCID IDs in publication and repository workflows; and by Tom Demeranville (ORCID) on how ORCID currently supports the collection of authenticated ORCID IDs for all authors in publication workflows. Problem statements around ORCID collection were then set out in presentations by:

- Catriona MacCallum (Wiley/OASPA) presenting a publisher viewpoint
- Philip Reimann (Scholar One / Web of Science) presenting the view from a manuscript tracking system provider
- Rory McNicholl (Co-Sector) presenting a view from a repository system provider
- Melissa Harrison (Europe PMC/EMBL-EBI) presenting the view from a disciplinary repository

The slides from each of the presentations accompany these notes. Tom Demeranville of ORCID set out the benefits of ORCID collection in publishing workflows, with an emphasis on the advantages offered by authenticating ORCID IDs and best practices for doing so.

The presentations were followed by a series of group exercises covering: How might authors and co-authors authenticate their ORCID IDs and what are the barriers to collection, authentication, and verification of ORCID IDs?

The group split into breakouts groups according to their area of expertise. They explored both routes to meeting the policy requirements highlighting actions required either from a submitting or depositing author, or from a co-author. For reference, route 1 involves open access publication in a journal, and route 2 involves making the authors accepted manuscript (AAM) or a copy of the journal's 'version of record' available via a suitable subject or institutional repository.

Route 1:

Workflow outline:

1. Corresponding/submitting author signs in using ORCID single sign-on
2. Submission forms are pre-populated using data from the manuscript and from ORCID record plus metadata from other PID registries (for affiliations, funding etc.) with a particular emphasis on ORCID IDs associated with UKRI grants
3. Any additional co-authors are added to the metadata, with names, affiliations and emails added
4. Emails requesting authentication of ORCID IDs sent to co-authors

Barriers and dependencies:

- Note that the 'corresponding' author and 'submitting' author may not be the same person
- Not all authors have or use their ORCID ID
- ORCID sign-on and authentication are not consistently used across publishers or journals
- Workflows to populate lists of co-authors from grant metadata or manuscript documents etc. are inconsistent and, in some places, do not yet exist
- Requires comprehensive list of researchers to be attached to grant IDs, with consistent updates throughout the life of a grant as individuals leave or join funded projects
- Requires all sources to record and share authentication status of IDs at each step
- If relying on authenticated IDs in grant metadata, a verification step would still be required
- Requires additional consideration of how researchers supported by core funding could be supported in meeting these policy requirements, as they will not be linked to a specific grant award
- Not all journals request/require co-author email addresses and have not set up process to contact them to validate ORCID ID

Route 2:

Workflow outline:

1. Authenticated ORCID IDs and associated institutional affiliations are collected by publishers and shared alongside article metadata at the point of acceptance
2. Notification of an accepted article is sent to relevant repository(ies) by a third-party service (Jisc Publications Router, OA switchboard, ORCID auto-update and notifications etc.)
3. The author provides a copy of the Authors Accepted manuscript (AAM) either on their own or in response to a prompt from repository staff
4. Repository staff verify connections between ORCIDs, institutional affiliations, funding etc. and publish suitable copy of the article (AAM or VOR if permitted)

Barriers and dependencies:

- Development work for repositories requires resources that are often not available at institutions
- Requires comprehensive list of researchers to be attached to grant IDs, with consistent updates throughout the life of a grant as individuals leave or join funded projects
- Requires consistent registration of DOIs at the point of article acceptance by publishers
- A central clearing house (described as 'AAM central' or a 'super-charged Router' by attendees) would simplify processes and improve coverage
- Requires all sources to record and share authentication status of IDs at each step
- Requires processes to manage duplicates at scale, with downstream effects (such as multiple entries for each output on individuals' ORCID records) mitigated

Group discussion and recommendations

The closing discussion focused on tangible actions for all stakeholders, and was directed towards questions such as:

- What can be catalysed/accelerated/amplified?
- What evidence is there for the scale of issues discussed?
- How to evidence the nature, extent, and variety of current journal practices around ORCID collection and authentication? This needs to take account of current and planned functionality of systems and opt-in levels.
- Could the group suggest evidence or indications of risks if actions are not taken on this issue?

Key actions recommended included **exploring the feasibility of a 'super-charged Publications Router' or a similar AAM/VOR feed**. The group felt that this could also be delivered via a system that would be something like Europe PMC for all UKRI-funded outputs

There is a critical dependency on UKRI registering Grant DOIs, and sector agreement on mechanisms to update those grant DOIs. These would also help to furnish evidence of real-world benefits to help motivate researchers to use their ORCID and connect across systems. This should be linked to a programme of advocacy across funding organisations to generate greater consistency and share potential costs.

Has UKRI gone as far as it can with regard to mandating ORCID use by authors? It likely has as it cannot mandate beyond UKRI funded authors, but it can add **more guidance for publishers, editors and researchers about the benefits and practicalities of using ORCIDs** and the value of doing this for all authors.

There is a clear need to **spell out what authentication means and the value of doing it**, as, for example, some publishers don't understand the SSO connection with having an authenticated ORCID (or what the implications are for having an authenticated ORCID). The group recommended that **ORCID/UKRI collaborate on a communications push on authentication**, including messaging for publishers around 'this is what you can do to help your authors be policy compliant'.

UKRI should clarify terminology around terms such as 'submitting author' or 'corresponding author'. It doesn't matter what the author role is, if they're UKRI funded their ORCID needs to be

included but publishers do need better clarity around what those roles are, and how the policy cuts across these categories.

Several options were discussed for ways to improve the coverage and sharing of authenticated ORCID IDs:

- If a submissions system detected a UKRI grant plus an associated ORCID ID it could trigger an email
- Grant submission systems will have ORCIDs associated with grants. Funders could submit additional information to Crossref. Currently there is additional functionality (post-hoc) around retractions, but could post-publication assertions be extended?

How much ambiguity/wiggle room/tiers of compliance can be created? The community would benefit from **clear criteria for what is compliant or not and having a time frame before full enforcement**. Repository managers need to be informed in advance. The policy should go in the direction of full compliance without creating huge disruption in the meantime, for example by accepting 'best efforts' narrative for repositories or by clarifying some exceptions, possibly similar to those allowable for REF, such as circumstances not fully in the researchers' control.

UKRI could **provide clear direction to researchers to publish where there are clear indicators of efforts towards compliance**. This would incentivise the publishers to support researchers.

A11: Summary of UKRI Virtual Round Table on PIDs for Funders and Research Performing Organisations

Recommendation

This virtual round table focused on MoreBrains' recommendation that UKRI: select ROR as the PID of choice for organisations and strongly recommend or mandate its use; work with funders to facilitate the transition to ROR IDs; work with publishers, repositories, and ROR to promote best practice; and provide specific instructions on what metadata fields need to be included and how they should be formatted.

Meeting date

October 3, 2023

Attendees

- Fiona Carr (ALPSP/Ringgold)
- Dominique Capostagno (NIHR)
- Tom Hibbard (Publishers Association/Sage)
- Hannah Hope (Wellcome)
- Hylke Koers (STM Association)
- Maria Gould (ROR)
- Catriona MacCallum (OASPA/Hindawi/Wiley)
- Valerie McCutcheon (ARMA/University of Glasgow)
- Fabienne Michaud (Crossref)

Facilitators

- Sara Ball (UKRI)
- Josh Brown (MoreBrains)
- Fiona Murphy (MoreBrains)

Notes

NB: The views summarised are those of participants and do not necessarily reflect the views, priorities and policies of UKRI.

Following this presentation by Sara Ball, Josh Brown, Fabienne Michaud, and Maria Gould, and after confirming that the scope of this element of the policy is journal articles, the main areas of discussion were:

- This recommendation is aiming to smooth a transition that is happening anyway
- There is a challenge for Crossref and repository data. The number of organisation PIDs being consistently used is pretty low. This situation is even messier for PID usage in repositories. A small number of repositories used in the analysis show good levels of metadata, but MB were only able to include about half the relevant repositories in the analysis
- There are plans to merge the Open Funder Registry into ROR, although there is no exact timeline for this, and Open Funder Registry will continue to be available until at least the end of 2024. The two entities contain different information so when merged should provide a measurably superior, functional PID. The transition is already underway on various fronts, from incorporating the data into ROR to the technical implementation on the Crossref side to support ROR IDs for funders. Users are already switching to ROR or making plans to switch
- Crossref and ROR are working closely together - check their respective blog sites. They are conducting a consultation with a small group of funders and planning to work with ScholarOne and Editorial Manager
- ROR supports relationships and hierarchies, as well as crosswalks. It can address many different use cases. See the case studies on the website. It is already supporting use cases for identifying funders. There will be a new organisation type “funder” available within the new schema, forthcoming in early 2024, which is currently in beta
- ROR use cases for funders and funding information - identifying funders, tracking outputs, compliance and discoverability of awards and systems
- Capturing funding information and including it in a structured form in publication metadata is essential for downstream tracking and reporting. Funders can also contribute by registering Grant IDs and including ROR in grant metadata. Supporting interoperability between these and making it easier to populate funding acknowledgements which then reach the final metadata (and published article) would be a big step
- The new version of the schema will enable funders to be distinguished from other research organisations. The funder type, e.g. non-profit, government, etc. can also be provided, and the field will be repeatable so that the metadata can distinguish between the same organisation acting as both funder and as research producer
- Discussion points
    - Pathways can be very different from funder and institutional perspectives

    - Try to make it as easy and efficient as possible for organisations to make the right decisions for them. This will support researchers, and also allow UKRI to monitor
- It was suspected there are instances where work should be labelled as “unfunded” but that is not happening. Research managers can work with the community on that, and the UKRI is conducting a separate relevant project, Monitoring and Evaluation Framework
    - A sub use case within this, would be that where the institution is, ipso facto, the research funder, then that organisation’s ROR should be included within the record
- It would be desirable for research organisations to be able to analyse/report using their own ROR
    - The OA policy encourages the use of organisation PIDs for author affiliations in articles. In line with the UK PID strategy, ideally the ROR should be included, although the system should be set up so that people would not need to know the ROR’s actual details (i.e. the system should be automatically pulling in PIDs via type-ahead or a similar mechanism)
- Going forward, if funders hold individual references, repository managers would be able to link to that data and update the repositories’ own records
- A wider question about the most efficient way forward would be if consistent tools for looking up funders and embedding information (e.g. something like the typeahead functionality provided when entering an affiliation in an ORCID record) could be made available
    - It was thought that this should be relatively easy to accomplish with the right lead time

What is a realistic timeline and what are the barriers?

Key barriers include adoption and inconsistency:

- Large publishers gather funding at submission but don’t require funding statements in the articles. Funding statements in the version of record would make it easier to make these statements machine readable and provide human readable acknowledgement of funders
- Infrastructure in general is hard for small and Diamond publishers. RORs are free, which could be a key factor when seeking to engage smaller publishers. This could help plot a route towards better metadata standards
- The submission process is another key barrier. If there are no key integrations at the point of submission, then that conversation needs to be happening now, e.g. ScholarOne is used for many journals published by the larger publishers and still needs to integrate ROR
- There are plans to integrate ROR with ScholarOne and Editorial Manager. This may take a while, but it is definitely going to happen. Meanwhile, Funder ID will be deprecated
    - The large submission systems (and other platforms) really need to hear from their customers (mainly publishers) that, e.g., ROR is needed/important
- Crossref organised a roundtable in June, where it was discovered that one organisation had commissioned an automated tool to extract funder information
- A big value add is that ROR allows the building of a mapping table to do crosswalks. It is also possible to search on any funder ID in ROR. This would also have the advantage of not forcing organisations to change what they’re already doing
- The larger publishers have already invested a considerable amount of money in development work for their systems which incorporate Ringgold. So they would need to be

encouraged to use ROR alongside Ringgold. A key rationale could be that RORs (unlike Ringgolds) can be included in metadata deposited in Crossref that is then made publicly available for the services that harvest from Crossref
- ROR provides an open (CC0) dataset so it cannot make commercial data available in its own records. This needs to be remembered when considering what sorts of mapping tools or services might be feasible
    - Ringgold has a relationship with ISNI which is open. So it can be used without cutting across licencing requirements, although this is not common or straightforward
    - ROR includes crosswalks to ISNI (though not vice versa)

Next steps:

- Getting institutions as well as funders (UKRI) to mandate funder and institutional RORs as a requirement for articles (even if phased or over an agreed time period) would make the biggest difference to adoption (although they tend to be doing this retrospectively through custom TDM rather than implementing into processes) Encouraging publishers to take a strong stance would also help.
- Funders and institutions could also encourage/require the inclusion of machine-readable funding statements in the VoR
- Working with OJS would help particularly support publishing activities for high volumes of users
- There could be an analogous piece of work required with the repository platforms to align workflows and enable that information to be entered into systems/metadata
- OA switchboard have been attempting to extract data from the manuscript submission systems, which has not been easy. Service providers need to be involved in this conversation
    - The multi stakeholder approach will help - if publishers request/offer the information in standard form that will make a big difference
- Workarounds need to be considered as part of these recommendations
- How can we create incentives to move forward? Organise further community-wide discussions and consistency in the asks

Useful links and resources

- Crossref blogpost referenced in the presentation
- ROR guidance on institutional lookups
- ROR mapping table for crosswalks
- An example of organisations working together to uncover hidden information

A12: Summary of UKRI Virtual Meeting on Preservation Location

Recommendation

This virtual meeting focused on MoreBrains' recommendation that UKRI work with CLOCKSS, Portico, etc. to develop mechanisms of cross-checking preservation locations that do not involve encoding this metadata at the article level.

Meeting date

September 19, 2023

Attendees

- William Kilbride (Digital Preservation Coalition)
- Alicia Wise (CLOCKSS)
- Dongqing Xie (Portico)

Facilitators

- Sara Ball (UKRI)
- Josh Brown (MoreBrains)
- Alice Meadows (MoreBrains)

Notes

NB: The views summarised are those of participants and do not necessarily reflect the views, priorities and policies of UKRI.

Following this presentation by Sara Ball of UKRI and Josh Brown of MoreBrains, the floor was opened for questions and clarifications. The main point raised was to confirm that the scope of this element of the policy is journal articles. In addition, following the meeting a definition of digital preservation (taken from the Digital Preservation Handbook) was added: "the series of managed activities necessary to ensure continued access to digital materials for as long as necessary… beyond the limits of media degradation, technical obsolescence or organisational change…".

How can the MoreBrains recommendation be delivered efficiently?

- During discussion of this question, the group considered who needs to be involved, what UKRI can do to support it, what communications or other interventions are needed, and what is a plausible timeline
- The participants noted that there are differences between the long-term archiving responsibilities of preservation organisations like CLOCKSS and Portico, and those of repositories which provide access to AAMs — and that attainable but meaningful expectations should be set for each
- Good practice is for articles to be preserved in three proper archives
- Legal deposit libraries were discussed as a preservation location, however, not all of them surface information about their e-holdings. For example, the British Library does not, while

France and Spain do (via The Keepers Registry, hosted by ISSN — discussed further below). Note, there is an existing metadata schema for this
- The Keepers Registry (KR), was discussed in some detail. Any archive with digital holdings can report them to KR; an API service for KR is currently being developed. There is no clear policy information for inclusion of preservation services but most participating organisations - including CLOCKSS, Portico, and some national libraries - are accredited archives and transparently document their active practices. Some, but not all, organisations that participate in KR are accredited
- A proof-of-concept project in the book space, led by CLOCKSS, could perhaps be adapted for journals. Participants include EDItEUR and OCLC, and they are testing adding preservation information to ONIX feeds, with verification being provided by preservation services. All data will then be embedded into OCLC's cataloguing records
- Putting together clear and concise guidance on how to comply with the UKRI policy will be critical
- Repositories were discussed and participants noted that, while they probably do not themselves provide preservation, their institutions almost certainly will (or should) and so other reps from those institutions may need to be brought into these discussions. Many institutions focus on preservation of the final version of record via CLOCKSS and Portico
- Three preservation scenarios for preservation were identified: 1) it doesn't exist, 2) it does exist and is documented, 3) it does exist but is undocumented/hard to find (3) may perhaps be the most common, exacerbated by the fact that there is a great deal of inconsistency and a lot of change in the repository community. While there are components of good practice in place at many repositories, they don't add up to good practice overall
- Again, it would be helpful to share examples of good practice, cooperation, success, and the Archaeology Data Service was noted as an exemplar
- DPC's Rapid Assessment Model could also be useful — it contains lots of supporting documentation on digital preservation
- Participants noted that there is a risk of duplication of effort by focusing on preservation (primarily of AAMs) in repositories as it's usually a requirement in library contracts that publishers take steps to preserve the VoR in accredited archives. For example, https://liblicense.crl.edu/resources/digital-preservation/
    - Note - CLOCKSS and Portico agreements make VoR of every article they preserve openly available following a trigger event
    - Note - it would be worth talking to PubMed Central about their preservation approach as an exemplar for other subject repositories. To be indexed in PubMed, journals must be archived with CLOCKSS or Portico
- Repositories typically assume that, for multi-authored papers, each repository needs to have its own copy, which leads to a risk of multiple DOIs for the same AAM

Potential UKRI actions/compliance options

- Work with KR to develop a formal policy, evaluate current accreditation status of organisations, etc
- To demonstrate compliance organisations would have more than one preservation location for all articles, which would be registered with KR. UKRI could conduct occasional spot checks to verify compliance

    - KR could provide a direct feed to UKRI and/or CLOCKSS, Portico, etc can also do so
- Mandate preservation locations to take the assessment, set targets for compliance, and check on progress Require organisations to share/publish a RAM assessment every X years
- Promote the fact that the UKRI policy only requires an AAM to be hosted on one repository

Links and resources

- Accreditation scheme for digital archives
- ISO 16363
- Survey of preservation capability of NPLD using a basis of Core Trust Seal
- Digital Preservation Coalition resources

A13: Summary of UKRI Virtual Round Table on Project PIDs

Recommendation

This virtual round table focused on MoreBrains recommendation that UKRI: works with RAiD, Crossref, publishers, and repositories to support RAiD as it matures into a project identifier that is fit for purpose; engages in communications to help the publishing, institutional, and researcher communities understand what a project is and how projects differ from grants; and provides specific instructions on what metadata fields need to be included and how they should be formatted once project IDs are sufficiently mature.

Meeting date

13 October 2023

Attendees

- Christopher Brown (Jisc)
- Matthew Cannon (Publishers Association/Taylor & Francis)
- Melissa Harrison (EBI)
- Ginny Hendricks (Crossref)
- Catriona MacCallum (OASPA/Hindawi)
- George Macgregor (UKCoRR/University of Glasgow)
- Valerie McCutcheon (ARMA/University of Glasgow)
- Adam Vials Moore (Jisc)
- Alok Pendse (Publishers Association/Elsevier)
- Shawn Ross (ARDC)
- Ben Ryan (UKRI)
- Natasha Simons (ARDC)

Facilitators

- Sara Ball (UKRI)
- Josh Brown (MoreBrains)
- Phill Jones (MoreBrains)

Notes

NB: The views summarised are those of participants and do not necessarily reflect the views, priorities and policies of UKRI.

After presentations from Sara Ball (UKRI) and Phill Jones (MoreBrains) explaining the context of the UKRI Open Access policy and the MoreBrains landscape analysis of sector readiness to meet the policy requirements, the group heard from Shawn Ross (ARDC) on the current status of development and adoption of the Research Activity Identifier (RAiD) for projects, and from Christopher Brown (Jisc) on a proposal to establish a UK RAiD registration agency (RA).

The group discussed the benefits of project PIDs, and shared open questions or considerations on the topic as follows:

- RAiD updates and information:
    - FAIRCORE4EOSC project will see RAiD integrated into the European Open Science Cloud
    - RAiD has applied to become a DOI RA, to use DOIs as the identifier for the RAiD record
    - The organisation registering the RAiD will be responsible for maintaining/updating the RAiD until the end date is added to the record
    - In active talks with potential RAs in Europe, North America, and China. Some discussions with individual organisations in South America, seeing how under-served geographies can be supported, possibly directly by ARDC
    - Need to ensure that RAiDs can be handed off to other RAs. For the UK, Jisc may be the RA
    - George Macgregor noted that Rioxx v 3.0 will be using RAiD, as per rioxxterms:project: https://rioxx.net/profiles/v3-0-draft-2/#rioxxterms:project
        - NB this is distinct from rioxxterms:grant: https://rioxx.net/profiles/v3-0-draft-2/#rioxxterms:grant
- RAiDs could be collected during article submission and included in the metadata for the article. Crossref would need to extend their schema to support RAiDs. This would also bring a need to figure out the relationship types, and who is stewarding which part of the record. Grants can also include project titles etc, not PIDs yet. Need to think about the workflows. Discussions at the DOI Foundation about how to link DOIs and RAiDs
- Shawn Ross anticipates integrations with ORCiD, DataCite, Crossref, etc. to ensure bi-directional relationships (you can already add the other PIDs to RAiD)
- The group discussed ways to make it easy to share and input RAiDs across systems with the goal of avoiding researchers having to enter the same data more than once
    - If the data is going into a system somewhere already, ideally it would get into RAiD records via the API
    - Want to have lookups etc. e.g. similar to the way that articles can be surfaced using author ORCID IDs
    - Better semantics in metadata to accommodate RAiD would be preferable, but nothing to stop a dumb relational link to a RAiD until then, and enabling agents to negotiate the link

    - Many researchers are already linking outputs to projects in OSF and Zenodo. Mirroring that easy upload/drag and drop would be good
- RAiD as a synchronisation tool (rather than as an additional layer of admin). Will need to tie in with existing internal project IDs
- Ideally a RAiD would exist before grant (etc.) application so information could be imported from RAiD to support the grant application. There are often a number of unsuccessful grant applications before a successful one. RAiD would enable the re-use of the data collected in that process between organisations
- RAID could help universities to discover previously unseen impact (including but not limited to citations) for particular projects, especially those funded by the university without explicit grant funding
- Before considering mandating RAiD, need to ensure there is a consistent and reliable level of adoption and coverage (NB: there are no current plans to mandate the use of RAiDs, it is is proposed as a project ID in line with the UK national PID strategy)
- Need to clarify relationships between RAiD and other initiatives that might 'resemble' a project PID, such as ROCrates (which is not a project ID, a package that ties a dataset to metadata about that dataset in one object) or other PIDs which are being used for projects, such as DOIs which are being used by Zenodo and OSF for project affiliations
- RAiD is an opportunity to pull together all the code and data that aren't already open and in a repository, making it much easier to find people's outputs. Important to note the difference between code re-use and the publications they are linked to. This is aligned with everyone's open research policies, from UNESCO on down and could be an easier way to pull all these things together
- Some universities don't want researchers anywhere near RAiDs - they want research support staff to maintain the record; others only want researchers to do it. There will be a need to support both pathways, especially for those organisations that can't afford a research information system
- Important to clarify the distinction between projects and grants
- Worth noting that block grants are often used to support things that were not planned at the time of award, projects will need to be able to connect a RAiD with a specific grant, programme etc. both after or before creation
- RAiD could be useful to encourage uptake and adoption of PIDs and metadata in the Humanities and Social Sciences as a means to track the impact and outputs of projects and research outputs that do not have traditional funding via grants

Useful links and resources

- RAiD API documentation
- Project metadata in DataCite

A14: Summary of UKRI Virtual Round Table on Self-archiving Policies

Recommendation

This virtual round table focused on MoreBrains' recommendation to use existing practices and workflows to cross check publisher self-archiving policies through SHERPA RoMEO, as stated in the UKRI OA policy.

Meeting date

28 September 2023

Attendees

- Jane Anders (Jisc)
- Imogen Batt (Publishers Association/Springer Nature)
- Andrew Beeken (Jisc)
- Ian Burgess (Publishers Association/Wolters Kluwer)
- Melanie Heeley (Jisc)
- Anton Heimann (Aries/Elsevier)
- Anna Jester (eJournal Press/Wiley)
- Catriona MacCallum (OASPA/Hindawi/Wiley)

Facilitators

- Sara Ball (UKRI)
- Phill Jones (MoreBrains)
- Fiona Murphy (MoreBrains)

Notes

NB: The views summarised are those of participants and do not necessarily reflect the views, priorities and policies of UKRI.

Following presentations by Sara Ball, Phill Jones, and Jane Anders, and after confirming that the scope of this element of the policy is journal articles, the main areas of discussion were:

- Sherpa services will be the new name for a suite of already existing services albeit with a new interface:
    - The Romeo dataset shows journal policies on OA.
    - Juliet shows funder policies on OA.
    - There is a joint compliance engine, (formerly known as FACT)
        - Shows overlap between funder and publisher policies so that researchers can see who they can and can't publish with
    - The OpenDOAR registry of open repositories.
    - The service suite includes information about what version should be self-archived, as well as a glossary of terms

- The key use case: The challenge for researchers: Funders have policies on lots of things and not necessarily consistent across funders. Researchers might be funded by multiple funders. Trying to figure out where you can and cannot publish can be a challenge.
- The plan is to retire these various names and just use the generic "Sherpa" (https://beta.sherpa.ac.uk).
- According to Jisc, approximately 91% of journal articles that are UKRI funded are made OA. The new system being developed will be able to provide analysis and reports more easily that is currently the case. There were no usage figures available, but Jisc's impression is that the Sherpa services are used heavily by university staff to advise researchers.
- Neither Crossref nor DataCite schemas contain self-archiving fields for individual article records. The required information is about the journal titles, not the publications themselves.
- Publishers are frustrated with having to contend with updates and policy changes. Born-OA publishers tend to have very consistent OA policies, while legacy publishers' policy mix is more complex (there's a more diverse mix of journal business models plus society-owned titles and sometimes even different policies per article type within the same journal).
- There is a need for UKRI to explain the rationale for the use of Sherpa services as it's currently unclear to publishers. This lack of clarity hinders engagement.
    - The answer is that repository teams and librarians use it to advise researchers on where they can publish based on their funders' policies
- The Coalition S Journal checker tool appears to publishers to do the same thing. Publishers would prefer a single mechanism and are concerned about supporting multiple platforms
- Many smaller publishers are probably not aware of Sherpa Romeo/Juliette and so are not using it, although the link to the DOAJ may be a way for them to have come across the Sherpa services.
- EJPress, which is part of Wiley Partner Solutions and provides software as a service for journals and societies, includes assorted functionality to pass along requirements and make authors aware of policies provided there is adequate clarity, and their journal/society/publisher customers request it.
- Publishers are most likely to use the service to look up other publishers' policies. Registering policies is valuable, but it is hard work that should not be under-estimated in terms of the challenges to keep records up to date.
- Currently publishers update their policy records via email, but there can be a long time lag between the instigation of the new policy and the update of Sherpa.
    - Policies need to be deconstructed and characterised by Sherpa, which is time-consuming.

Potential next steps

- Convene a discussion around the standardisation of language around self-archiving policies, particularly around definitions of article/manuscript types.
- Development of machine-readable policies that can be computationally ingested and compared systematically
    - NISO may need to be engaged to help with standards (there are NISO guidelines from 2008 that could be updated)

- Update the ways that policies are updated as without a standard it is hard to ingest the policy parameters. With a machine-readable standard, Sherpa could be given feed addresses for the policies and ingest automatically.
- The policies could be modular. It would be good to get away from a model that depends on a few people remembering to provide the information. Sherpa could create a template that can be used by different policies?
- UKRI to review the landscape as there are other organisations interested in this space, such as COKI (Curtin University's Open Knowledge Initiative), https://www.journalobservatory.org/, Journal Checker Tool, COAlition S, OA Switchboard. The goal is to coordinate in order to reduce channel confusion for publishers and make it clearer what is needed.
- In the short term, work with one or more smaller publishers to develop workflows and case studies on how to pass policy information efficiently to everywhere it needs to go. Need to demonstrate how time can be saved and levels of service for authors improved
    - eJournalPress may be interested in collaborating
- Sherpa FACT (Funders & Authors Compliance Tool) provides funder policies, compares journal policies, and advises whether grant awardee can publish on that basis. The Sherpa redevelopment project is modernising and making this service more flexible.
- UKRI to conduct more evidence gathering on what information is needed and how it should be reported.
- Work with smaller publishers and vendors to develop a case study as a model for other publishers to follow.
- Develop collateral aimed at publishers on why they should participate in and update Sherpa. Include reasons such as being good citizens, supporting needs of institutional customers worldwide; and supporting researchers worldwide.
- Encourage the concept of providing policies as a service. Some metrics would be helpful.
- Engage with publishers and vendors on structured formats and workflows for easy reporting, such as JSON or XML.
- Develop a roadmap or plan for a longer-term transition towards standardised, machine-readable policies.

Outstanding questions

- The effect of rights retention policies.
- The effect of potential changes in publisher business models, e.g. charges for manuscript processing when rights retention language is applied.

A15: Summary of UKRI virtual workshop on Versions of Record, Authors Accepted Manuscripts, and Repositories

Recommendation

This virtual workshop focused on MoreBrains' recommendation that UKRI: work with the repository community to promote best practice in version recording; promote the registration of DOIs or other PIDs for content; promote the use of RIOXX or the DataCite schema, giving repositories freedom to choose the approach that works best for them; provide specific instructions

on what metadata fields need to be included and how they should be formatted; and work with Crossref and Publications Router to develop ways to supplement repository metadata.

Meeting date

16 October 2023

Attendees

- Steve Byford (Jisc)
- Todd Carpenter (NISO)
- Patricia Feeney (Crossref)
- Ginny Hendricks (Crossref)
- George Macgregor (UKCoRR/(University of Glasgow)
- Jeroen Sondervan (NWO)

Facilitators

- Sara Ball (UKRI)
- Josh Brown (MoreBrains)
- Phill Jones (MoreBrains)

Notes

NB: The views summarised are those of participants and do not necessarily reflect the views, priorities and policies of UKRI.

The workshop started with this presentation by Sara Ball and Phill Jones as background to the project, clarifications about how the data in the original report was compiled, and a verbal update (no slides) from Todd Carpenter.

- The problem as defined is that in order to be compliant with the UKRI OA policy, if an author wishes to comply using Route 2 (Authors Accepted Manuscript - AAM - placed in a suitable repository) the repository record must contain the resolvable PID (in practice, usually a DOI) of the published Version of Record (VoR), with the version held in the repository clearly identified and linked, via the PID, to the VoR
    - Currently, repository metadata and API endpoints are inconsistent making it impossible to systematically aggregate the data or monitor compliance
    - For the MoreBrains analysis, CORE was able to measure compliance for about 50% of institutional repositories in the UK (those that make use of the RIOXX application profile)
    - Around 75% of records that could be analysed included a DOI. However, it is unclear if those DOIs were for the Version of Record or were DOIs allocated to the repository version itself
- Todd Carpenter gave an overview of the history, rationale and current status of NISO's work on article versions. NISO published the journal article versions recommendations in 2008, in collaboration with ALPSP. It specifies the versions that a manuscript goes through on the way to final publication

- The original standard is publisher-centric because it was deemed that each stage in publication adds value. Stages were:
  - Original submitted manuscript
  - Manuscript under review
  - Accepted manuscript
  - Accepted proof
  - Version of record
  - Corrected VoR
  - Enhanced VoR
- The standard doesn't include pre-prints, which is a major driver for the current work on revising the recommendations. This work isn't finalised, it's still under discussion. There is also a greater focus on how terminology is used outside of publishing. The current list of terminology will not likely be replaced, but there will be more context

Group discussion centred on metadata adoption and completeness challenges. Issues raised included:

- RIOXX adoption is not high enough, or improving fast enough
- Links to all related outputs still need to be captured
- Need to bring clarity to the policy version requirements. Many in the repository community do not view pre-prints or AAMs in the way the policy does
- The more information that authors (or other depositors) can provide at the time of repository deposit, the easier subsequent matching to VoRs will be
- Few publishers contain links to other article versions in VoR in metadata, although despite resistance some years ago, connections to pre-prints are becoming more common

Group exercises

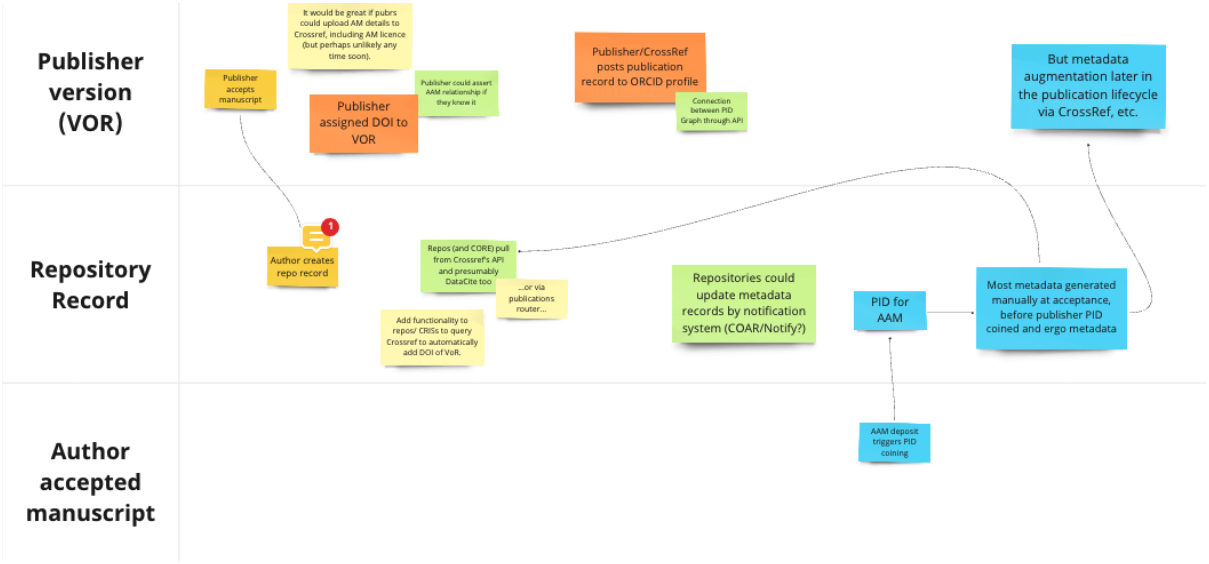
After the initial discussion, the group undertook two exercises using Miro. The board is here.

The first exercise explored the metadata needed to enable the identification of versions and the mapping of relationships between different versions (in this case, the AAM and the VoR). The group felt that a solid core of metadata must be provided to enable matching between versions for which key metadata may have changed (e.g. a changed title between acceptance and publication). To maximise the efficient linking of versions, the group recommended the collection and open availability of the following metadata elements:

- DOI (or other suitable PID for each version referred to in the metadata record
- ORCID IDs of as many authors as possible (NB: the policy requires all UKRI-funded authors to provide their ORCID ID)
- ROR IDs for author affiliations
- Funding acknowledgements, ideally with grant DOIs
- Title
- Journal Article Version per NISO standard
- Relational information between the repository version and all known other versions (e.g. pre-prints as well as VoRs)
- Dates of submission, acceptance and publication
- Publisher(s) and publication venue(s)

The goal of the second exercise was to map how the DOI of the VoR can get into the repository record and identify interventions to improve accuracy and timeliness.

Here is the workflow that the attendees came up with:



- At the point of manuscript acceptance, the researcher, or an administrator, creates a repository record
- In many cases, the repository record and associated metadata is generated before the DOI is known, meaning that it is impossible to add the VoR DOI at the point of creation. It therefore must be updated
- Publishers assign a DOI to the VoR
- In ideal cases, the repository pulls the DOI from a metadata registry that may be held by CORE, Crossref or DataCite, but this workflow is not well implemented across the sector and there is insufficient automation
- In addition, records of publications can be passed to the author's ORCID record from Crossref, if they have enabled auto-update. Repositories could use the ORCID record to enrich their own metadata, however that workflow is currently rarely implemented or automated
- There are notification systems (e.g. COAR/Notify) that could be used to update metadata in the repository

Interventions suggested

- Repositories should reach a consensus on what PID they want to use for articles and other entities. Many repositories use DOIs, but many choose not to, and use Handles, ARKs, or other identifiers
    - Consistency in PID usage would make matching and linking much easier
- Guidance for authors on workflows and digital scholarship is needed
- Publishers should collect AAM PIDs. This should be incentivized or automated
- Encourage earlier PID creation using "Pending Publication", which is supported by Crossref to increase the number of VoR DOIs that can be entered when repository records are created

- Support repositories to automate metadata enrichment via Crossref, ORCID, CORE, etc APIs
- Investment in open solutions for filling in metadata gaps is needed. For example, Crossref matches articles based on titles and authors, when a PID is not available.
- Investigate the use of the NISO Manuscript Exchange Common Approach (MECA) Recommended Practice, or other approaches, to automate information exchange between publishers and repositories
    - MECA enables the metadata that was already entered by the researcher into a repository or CRIS to be automatically transferred into publisher submission systems. The reverse is also possible.

Other observations and suggestions:

- Work is needed to standardise the use of DOIs in repositories so that it's clear which DOI refers to the VoR, and if the repository uses DOIs, which DOI refers to the repository version
- Researcher and repository engagement is needed to raise awareness and standards of digital scholarship
- Increasingly, scholarly works are multi-part objects. So further relational information will eventually be required between the components of repository records and their published VoRs. E.g. A repository record may include an article, a dataset, engagement materials, and multimedia content

Barriers

- Within many institutions, repositories have been retired and replaced by CRIS systems with repository modules. There is an associated lack of understanding and capability around digital scholarship
- Inconsistency in when PIDs are availability means that multiple workflows need to be supported
- Uncertainty and variance in PID usage and adoption, metadata schemas and API endpoints among repositories creates a confusing landscape that is difficult for funders and publishers to engage with

Useful links and resources

- NISO Manuscript Exchange Common Approach (MECA) Recommended Practice
- NISO Journal Article Versions (JAV)

Appendix B: Contributors to this project

Technical Requirements Project Group membership

Michael Ball, UKRI (MRC)
Sara Ball, UKRI
Christopher Brown, Jisc
Josh Brown, MoreBrains
Rachel Bruce, UKRI
Matt Buys, DataCite

Steve Byford, Jisc
Melissa Harrison, EuroPMC
Karen Jackson, Jisc
Phill Jones, MoreBrains
Petr Knoth, CORE
Catriona MacCallum, OASPA
George Macgregor, UKCORR/RIOXX
Valerie McCutcheon, ARMA
Lee Miller, Falmouth University
Ben Ryan, UKRI (EPSRC)
John Salter, University of Leeds
Wayne Sime, ALPSP
Steven Vidovic, University of Southampton
(Observers from the National Institute of Health Research)

Participants in community consultation and outreach events

Jane Anders, Jisc
Frederick Atherden, eLife
Imogen Batt, Springer Nature
Andrew Beeken, Jisc
Justin Bradley, University of Southampton
Christopher Brown, Jisc
Ian Burgess, Wolters Kluwer
Stacey Burke, APS
Matt Buys, DataCite
Steve Byford, Jisc
Matthew Cannon, Taylor & Francis
Dominique Capostagno, NIHR
Todd Carpenter, NISO
Fiona Carr, CCC
Alina Chowdhury, Jisc
Naomi Conforti, Clarivate
Fergus Crisham-Nimmo, OUP
M. Cruz, NWO (Dutch Research Council)
Michael Cunningham, Frontiers
Tom Demeranville, ORCID
Eleanor Dumbill, University of London
Tami Ezra, Clarivate
Ashley Farley, Gates Foundation
Patricia Feeney, Crossref
Rocio Gaudioso, Frontiers
Maria Gould, ROR
Mark Hahnel, Digital Science
Melissa Harrison, EBI
Mark Heaver, Taylor & Francis

Melanie Heeley, Jisc
Anton Heimann, Aries Systems
Ginny Hendricks, Crossref
Kelly Hetherington, Durham University
Tom Hibbard, Sage Publications
Matt Hodgkinson, UKRIO
Hannah Hope, Wellcome
Iain Hrynaszkiewicz, PLOS
Lisa Hulme, Clarivate
Azhar Hussain, Jisc
Anna Jester, eJournal Press
Peter Jones, OUP
Daniel Keirs, IOP Publishing
William Kilbride, Digital Preservation Coalition
Petr Knoth, The Open University
Hylke Koers, STM Solutions
Rachael Lammey, Crossref
Leslie Lansman, Springer Nature
Adam Leary, OUP
N. Lock, Elsevier
Catriona MacCallum, Hindawi
George Macgregor, University of Glasgow
Paloma Marín-Arraiza, ORCID
Agustina Martinez-Garcia, University of Cambridge
Valerie McCutcheon, University of Glasgow
Rory McNicholl, University of London
Kirsty Merrett, University of Bristol
Fabienne Michaud, Crossref
Adam Vials Moore, Jisc
Leila Moore, Wiley
Helen Nolan, NIHR
Richard O'Beirne, OUP
Sam Parker, Hindawi
Oksana Parylo, Frontiers
Lou Peck, The International Bunch
Alok Pendse, Elsevier
Ed Pentz, Crossref
Sam Perkins, Sage Publications
Simon Porter, Digital Science
Iratxe Puebla, DataCite
Philip Reimann, Clarivate
Nigel Robinson, Clarivate
Shawn Ross, ARDC
Ben Ryan, EPSRC UKRI
Lorianne Sarsfield, Sage Publications

A.K. Sehgal, Elsevier
Kathleen Shearer, COAR
Wayne Sime, ALPSP
Natasha Simons, ARDC
Sharon Skelton, FCDO
David Sommer, Kudos
Jeroen Sondervan, NWO (Dutch Research Council)
Rebecca Taylor-Grant, Taylor & Francis
M. Teperek, NWO (Dutch Research Council)
Diane Thomas, Clarivate
Inesa Thomsen, DHSC
Sandra Townsend, Wiley
Gonçalo Vargas, Frontiers
Nathan Westgarth, Aries Systems
Alicia Wise, CLOCKSS
Dongqing Xie, Ithaka