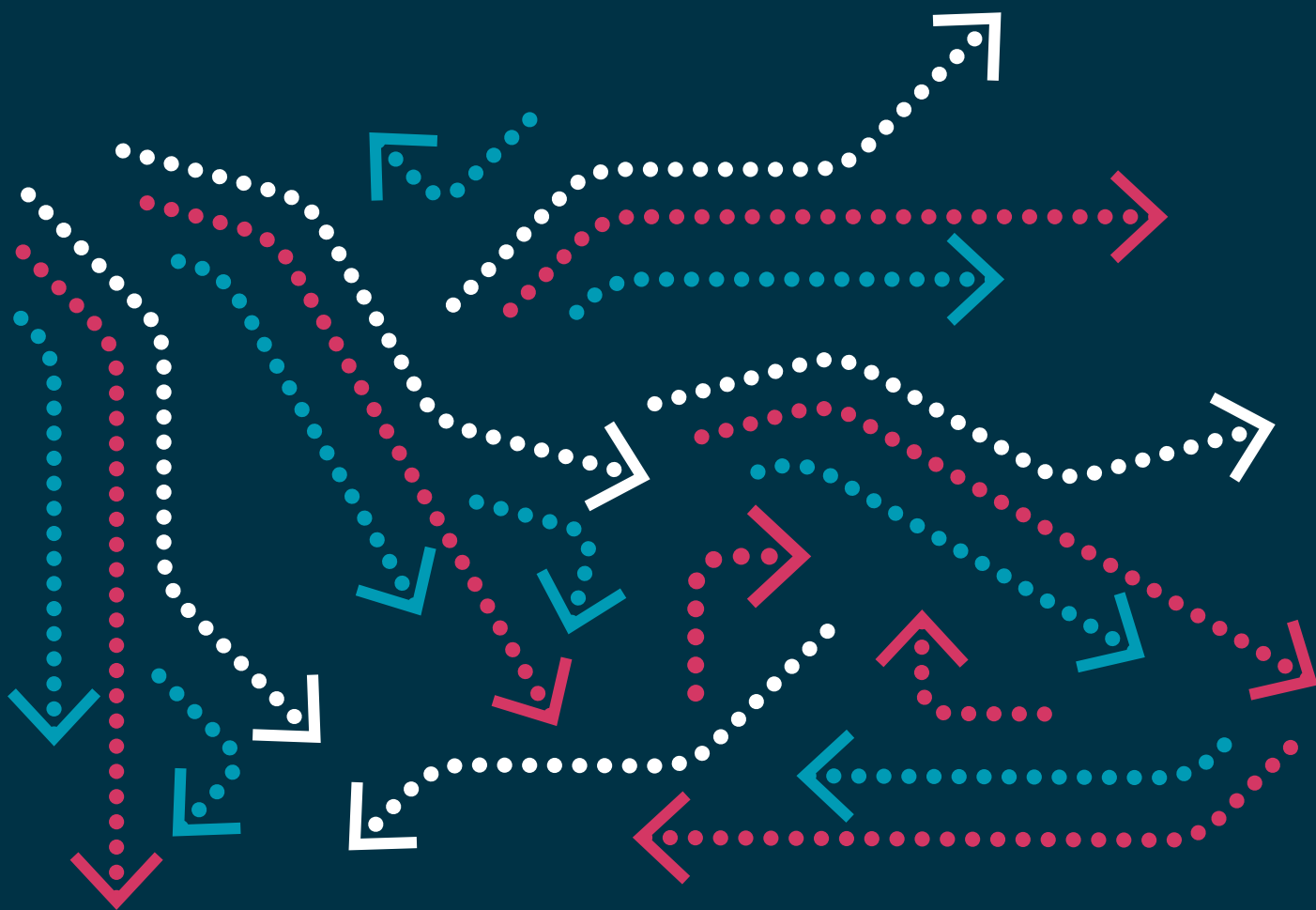


**UNIVERSITIES UK
OPEN ACCESS
COORDINATION GROUP**

JULY 2018

**OPEN ACCESS REPOSITORIES:
REPORT AND RECOMMENDATIONS**



Universities UK

OA REPOSITORIES WORKING GROUP

CHAIR: Anne Horn (University of Sheffield)

SECRETARIAT: David Ross (SAGE Publishing)

1. Introduction and background

As part of Professor Adam Tickell's advice given to former Minister for Universities and Science, Jo Johnson, on [open access to research publications](#) (2016), a recommendation was made that a Repositories Working Group (WG) be convened. This recommendation reflects the fact that repositories have been given a strategic role to play at an international level (e.g. H2020), national level (e.g. Higher Education Funding Council for England's (HEFCE) Research Excellence Framework (REF) policy), and institutional level.

There is a wide variety in the nature, roles, and functions of repositories, including, but not limited to:

- **Focus:** institutionally-based or discipline-based
- **Location:** UK, UK-mirrored, international
- **Deposit:** author, mediated, publisher
- **Access:** open, public, discoverable
- **Type of content:** peer-reviewed articles and monographs, reports, working papers, preprints and other grey literature, theses, dissertations
- **Versioning:** pre-prints, post-prints, final formatted versions, mixed
- **Embargos:** none, author-set, funder-defined, publisher-set
- **Rights:** clear/opaque, CC, others
- **Preservation status:** does the repository have a formal preservation function?

2. Objectives

The Repositories working group was tasked to investigate and seek to mitigate challenges around:

- a. Enabling fulfilment of funder and institutional mandates
- b. Easing deposit of relevant versions of research outputs
- c. Creating and updating high-quality metadata records through the lifecycle of a research output
- d. Integration of repositories into the wider scholarly communication landscape, including with identifier services
- e. Improving discoverability of content in repositories
- f. Effective aggregation of repository content to improve text and data mining and other actions requiring access to a corpus of literature
- g. Provision of reporting and auditing structures for funders and institutions
- h. Ensuring long-term preservation of repository content
- i. Ensuring repositories evolve to exploit technologies and practices to allow greater interoperability
- j. Ensuring clear expression of embargo and licencing information
- k. Exploring the costs associated with the storage of an article in a repository, as part of longer-term forecasting
- l. Ensure effective communication of requirements and mechanisms to the research community

The WG acknowledged in its consultations and discussions that there is considerable overlap in these objectives.

3. Membership

The WG brought together stakeholders from relevant organisations, broadly reflecting the membership profile of the Universities UK Open Access Coordination Group.

Academic representatives	<ul style="list-style-type: none">• Jo Fox (Durham University)
Jisc	<ul style="list-style-type: none">• John Kaye
Library representatives	<ul style="list-style-type: none">• Anne Horn (University of Sheffield)• David Prosser (RLUK)• Masud Khokhar (SCONUL)• Torsten Reimer (British Library)
Publisher representatives	<ul style="list-style-type: none">• David Ross, Secretary (SAGE)• Edward Wates (Wiley)• Malavika Legge (Portland Press/Biochemical Society)
Research Administrators	<ul style="list-style-type: none">• Simon Kerridge (University of Kent)• Michelle Double (University of Leeds)
Subject-based repository	<ul style="list-style-type: none">• Karen Vogtmann (University of Warwick)
UKCoRR (United Kingdom Council for Research Repositories)	<ul style="list-style-type: none">• Yvonne Budden (University of Warwick)
Funder Representative	<ul style="list-style-type: none">• Geraldine Clement-Stoneham (Medical Research Council)• Robert Kiley (Wellcome Trust)

The WG met on three occasions in person; 10 February 2017 (Wellcome Trust, London), 9 June 2017 (British Library, London), and 24 October 2017 (British Library, London). Between meetings, the sub-groups facilitated discussions among members in their areas of interest and expertise, as well as consultation with their respective communities.

4. Engagement

A survey of members of [UKCoRR](#) sought feedback from the higher education institution repository community around issues of concern. 47 institutions responded, ensuring that the WG formed a comprehensive view of where improvements are being sought. Respondents were also asked to nominate stakeholders best placed to take action. The responses have informed the WG's recommendations.

The Universities UK Open Access Repositories Group also co-hosted a workshop with the British Library on 28 July 2017 to explore the interest of higher education institution institutions in a national solution/shared approach to digital preservation of open access (OA) outputs. The workshop seemed timely given developments happening both at the British Library and Jisc. The British Library is in the process of enhancing its digital preservation capacities and looking at how it may expand its services to others, including

theses ([EThOS](#)) and OA publications. Jisc is piloting a [Research Data Shared Services Model \(RDSS\)](#) with a number of higher education institutions, and digital preservation is in the project scope. The workshop attendees heard more about these developments, as well as re-visiting earlier work undertaken by SHERPA Jisc in 2003 looking to the feasibility of preservation services.

5. Discussion

The richness of the repository landscape in the UK has been a key enabler in advancing progress towards OA in this country. [OpenDOAR](#), the Directory of Open Access Repositories, lists 256 UK repositories, marking the UK as the country with the second largest number of repositories in the world behind only the United States.

The development of institutional repositories has evolved significantly in the UK and internationally since their implementations in the early 2000s. The UK's policy for advancing OA for research outputs has been a key driver in this country, with evidence that the pace of adoption of OA by researchers has gained momentum between 2012 and 2016, as evidenced by the report [Monitoring the Transition to Open Access](#) (December 2017).

Many institutions have expanded their repository requirements addressing other strategic imperatives for their institutions, including the provision of a publishing platform for grey literature, and digital collections of original materials.

There is a vital spirit of collaboration among institutional repositories, with particular emphasis on linking up staff knowledge and expertise. This is most evident through the activities of UKCoRR, a professional membership organisation. In some instances, technological infrastructure is shared. The [White Rose Consortium](#), for example, provides a shared repository for the University of Leeds, University of Sheffield and University of York. Some organisations provide hosting for other universities, including the University of Edinburgh and the University of London Computer Centre.

The UK benefits from the services and the ongoing research and development undertaken by Jisc. Repository managers have relied on SHERPA services for many years for compliance with publisher policies. There has been progress for repositories with the [Jisc Publications Router](#) and systems' interoperability with the most commonly used repository platforms. To see the benefits of this investment, more publishers need to come on board. When publishers routinely share or 'push through' the full text for Author Accepted Manuscripts and the accompanying metadata, then readers, researchers and publishers will see efficiencies in workflows and improvement in discovery. In addition to [current content providers](#), a number of major publishers have signalled their intent to do this.

Progress has also been made with the adoption of persistent identifiers to improve efficiencies and aid discovery. ORCID iDs, DOIs and Funder IDs address issues around consistency and transparency, providing machine-readable links between the identities of researchers, their works and their affiliations, higher education institution or research organisation, funding council, and publisher.

Over 65 UK institutional repositories have adopted the [RIOXX](#) Metadata Application Profile which provides a mechanism to help institutional repositories comply with [Research Councils UK \(RCUK\) policy on OA](#).

The need for an agreed, community-wide solution for institutional identifiers is pressing, and requires a project champion to take it forward. The inability to disambiguate the author affiliation in a standardised way is a major obstacle to the automation of many solutions. Commercial systems, such as [RingGold](#), provide a part solution as do non-commercial systems such as [GRID](#), developed by Digital Science. Recently, members of the persistent identifiers community met to discuss setting up an [Organisation ID Registry](#) with international community governance.

Preservation of OA content is recognised as a need, however, the pressure on each higher education institution repository to find its solution is a key concern. Consultation with the OA community through UKCoRR, followed by a workshop hosted by the WG and the British Library has provided evidence of this. There is interest among the repository community in exploring the feasibility of a national solution with key stakeholders.

Researchers are also supported by a growing family of subject repositories. The [Wellcome Trust](#) – with support from 28 funders across Europe – has long supported [Europe PubMed Central](#), which provides access to over 4.3 million full-text biomedical research articles and over 30 million abstracts. [ArXiv](#), with more than 1.3 million e-prints, has a similarly high reputation, and is being joined by new preprint services; [bioRxiv](#) is one example.

JISC's CORE services provide opportunities for different stakeholders, including companies, researchers, and higher education institutions, to analyse and mine data aggregated from repositories; to implement APIs for improving workflows and developing new services; and to implement technologies that improve harvesting of and access to OA content. Since 2015, [Core's OA full-text content has grown to over 10 million](#). For repositories, technical capability at the local level and effective partnerships with suppliers of library discovery services, underpin the take up of some services.

A high and growing proportion of UK-authored publications are published under immediate OA publishing models, over and above strong global growth ([Monitoring the Transition to Open Access](#), December 2017). For the UK, this represents 29.5% of publications available immediately, with over 18% of this increase achieved between 2014 and 2016.

While much has been achieved, stakeholders representing higher education libraries and research administrators, publishers, learned societies and funders recognise benefits in improving our communications and reporting, systems, workflows, data elements, and training and support.

- For the **institutional repository community**, issues around technical support and skills, including integration with identifiers and other university systems, rank highly in terms of operational concerns. The lack of reporting facilities to meet institutional and funder requirements adds an unnecessary burden. The management effort associated with journal embargoes and tracking author accepted manuscripts are also commonly identified key contributors to inefficient processes. More refined metadata schemas were seen as essential.

Taking a longer view, there are concerns about the architecture and, therefore, sustainability of some of the aging repository platforms, particularly where technical support is not readily available.

- **Academic members of the WG** expressed concerns that centre not on the nature of repositories themselves, but on any procedures that may not take into account that researchers should be free to publish in the most appropriate outlets for their work; and that this work should continue to be recognised in relevant assessments such as the REF.

Researchers have asked that appropriate consideration be given to issues surrounding 'versioning' with many in favour of having only one version of a work available. There are tensions with regard to differences in HEFCE and RCUK policy and the requirement of research councils for manuscripts to be made available without restriction on non-commercial use.

Researchers still value the role that peer-review plays in ensuring academic standards, and though the increase in the use of pre-printing is welcomed, researchers in the humanities in particular have stressed that these should always be linked to the final, peer-reviewed version. Researchers would be opposed to proposals that result in any additional administrative burden on both individuals and institutions.

- From the **publisher community**, representatives identified the need to agree standard delivery destinations and protocols within a stable regulatory framework as the priority. Delivery protocols should adhere to industry standard methodologies based on API solutions, and adopt standard metadata schemas in relation to article version, copyright licensing, embargo period, and author taxonomies. A standardised approach as demonstrated by PubMed Central or the Jisc Router is to be encouraged. Investment in the required delivery solutions is dependent on there being a stable set of regulatory guidelines and principles.

Discovery, long-term sustainability, and preservation continue to be big challenges and there is a need to define services supporting open research that require a national approach.

Although content in institutional repositories is typically exposed to Google and Google Scholar, discovery of content held in repositories is still problematic. Even when this content is discovered, the user is often provided with access to a flat PDF file which has inherent restrictions; for example, it is not easy to access the underlying data, nor facilitate machine-to-machine access.

The higher education institution repository community has concerns about the sustainability of aging repository platforms and systems development work required to improve functionality and interoperability. Concomitantly, institutional repositories are under increasing pressure to find a preservation solution for which they are not funded nor have the technical expertise.

6. Recommendations

Repositories for the most part were initially set up to meet the needs of institutions and subject communities. They are now meeting a broad national need in support of OA and in so doing form an essential component of national research infrastructure.

Institutional repositories operate in a federated environment in the UK. Subject repositories are also features of the OA landscape, with the [Europe PMC](#) repository model endorsed by the WG as an example of best practice. [Jisc services supporting OA](#) are of value to higher education institutions. However, discovery, long-term sustainability and preservation continue to be big challenges. The feasibility of a national services model to address one or all of these elements is seen by many as the next step in this transition to OA, with the British Library and Jisc already sharing expertise in this space.

The WG has recommended a set of requirements for adoption by stakeholders to improve the current federated model of institutional repositories. Concomitantly, the WG has identified as an outcome of its consultations, support for a study into national solutions addressing the big challenges of discovery, sustainability, and preservation, with the aim to provide enriched services.

6.1. To ease deposit and improve communications for authors, greater take up of metadata schema as applied at different points in the lifecycle of a research output is required. Discoverability of content in repositories will also be improved including clarity around version of record.

It is recommended that

6.1.1 Licensing terms be clearly articulated by publishers in machine-readable form, including PDF and HTML renderings.

6.1.2 Both publisher's systems and repository platforms provide machine-readable metadata that includes [NISO/Crossmark](#) article version tags, licensing tags and embargo periods consistent [RIOXX](#).

6.1.3 Repositories require a 'project/funder' field as part of metadata provision for a research output. Publishers are advised to collect funder information, at the point of submission and release this with the article metadata. To support this, publishers and other stakeholders are encouraged to make use of the [Funder Registry \(CrossRef\)](#).

6.1.4 Institutional repositories aim to be fully compliant with [OpenAIRE](#) to allow automated OA reporting to research councils.

6.1.5 Publishers should allow manuscript deposit under license terms that facilitate text and data mining and allow academics to meet funder requirements.

6.1.6 Publishers, funders and research organisations are encouraged to actively engage with the [Organisation ID Registry](#) initiative.

6.2. To improve the efficiencies of workflows and integration of repositories into the wider scholarly communication landscape OA community-wide action is required.

It is recommended that

6.2.1 Publishers and institutional repositories fully implement the [JISC Publications Router](#), supported by Jisc-negotiated publisher agreements for metadata and accepted manuscripts.

6.2.2 Providers of a Current Research Information Systems (CRIS) or related solutions be required to fully integrate with the Jisc Publications Router.

6.2.3 Higher education institutions, funders and publishers integrate ORCID iDs into their systems and workflows.

6.2.4 Publishers share with authors and higher education institutions [Digital Object Identifiers \(DOIs\)](#) for research outputs at the point of acceptance, where feasible.

6.2.5 The work being done by [CASRAI](#) to establish standardised contributor role taxonomies, CRediT, be supported by all stakeholders.

6.3 To ensure long-term preservation of repository content.

It is recommended that

6.3.1 A study into the feasibility of a national preservation solution be undertaken, recognising that the British Library and Jisc are key stakeholders.

6.4 To improve the capacity of stakeholders across the OA community to exploit new technologies and to improve discovery and reporting mechanisms, effective working relationships, and the channels of communication required.

It is recommended that

6.4.1 Higher education institutions, Jisc, subject repositories and other stakeholders take forward as a high priority improvements in the user experience and user interfaces, leveraging relationships with commercial system providers and open source communities.

6.4.2 Higher education institutions ensure that there is appropriate capacity for managing and developing repositories, including training and support.

6.5 To improve discoverability, further work is needed to define services supporting open research that require a national approach.

It is recommended that

6.5.1 A study be conducted to explore the need for national repository solutions or 'hubs' for one or all of the big challenges – discoverability, sustainability and preservation. This study will consider costs and benefits, and ultimately seek to define the guiding principles and services that a national hub will provide for the benefit of higher education institutions, and the advancement of the UK government's ambitions around open science.

This publication has been produced by Universities UK, the representative organisation for the UK's universities. Founded in 1918, its mission is to be the voice of universities in the UK, providing high-quality leadership and support to its members to promote a successful and diverse higher education sector. With 136 members and offices in London, Cardiff (Universities Wales) and Edinburgh (Universities Scotland), it promotes the strength and success of UK universities nationally and internationally.

Woburn House, 20 Tavistock Square
London, WC1H 9HQ

Tel: +44 (0)20 7419 4111

Email: info@universitiesuk.ac.uk

Website: www.universitiesuk.ac.uk

Twitter: @UniversitiesUK

To download this publication, or for the full list of Universities UK publications, visit www.universitiesuk.ac.uk

July 2018

ISBN: 978-1-84036-395-1



Universities UK

