

Plan S and institutional repositories – a response from UKCORR

8 February 2019

The United Kingdom Council of Research Repositories¹ (UKCORR) represents about 600 library professionals and research administrators across 150 UK institutions working with and supporting open access repositories. Collectively, the members of UKCORR have significant experience in the implementation and development of national, funder and institutional open access policies.

We strongly support the overarching aims of Plan S for the immediate and free access to scholarly research. This has been a long time coming and UKCORR sees this as a welcome step in the ongoing move towards open research. However, as with most things, the devil is in the detail. In the UK, research organisations have been grappling with the open access policies of government and charity funders for many years, and while Plan S is a step in the right direction towards policy alignment, the requirements for Plan S compliance are not straightforward. We very much welcome the opportunity provided by cOAlition S to request clarity on particular points relating to institutional repositories in relation to Plan S.

For many researchers, especially those actively engaged in open research, overly bureaucratic restrictions on how their research should be made open access can be confusing. Take, for example, the open access policy for the Research Excellence Framework (REF) 2021. The REF policy's ambition was to "provide a set of minimum requirements for open access"² but it took four years of lobbying before preprint servers were recognised as a suitable open access source³. The open research landscape changes quickly, and policies developed today can unwittingly stifle tomorrow's innovations. We therefore hope that cOAlition S will learn from past missteps and develop very clear guidance for institutions and authors to follow.

Within the open research landscape, institutional repositories are essential for many authors, and the Plan S implementation guidance, in its current form, may marginalize repositories even further than present. We believe that open access repositories add to the diversity of open research platforms and should therefore be nurtured and supported by funding bodies to become an essential component of the scholarly communication landscape. Failure to do so will exclude a large proportion of 'APC poor' authors from their participation in scholarly discourse, as well as inhibit the growth and exposure of high-value grey research literature. And without clear leadership from cOAlition S to support routes to free gold open access, the association between gold open access and APCs will only grow. The strength of open research is in its diversity, and in many respects Plan S seeks to strictly define the way in which research should be shared.

UKCORR would like to ensure that the mistakes of the past are not repeated. The Plan S implementation guidance for open access repositories presents some serious and significant challenges for the repository community. The requirements are reproduced below:

"10.2 Requirements for Plan S compliant Open Access repositories:

¹ <https://www.ukcorr.org/>

² https://www.ref.ac.uk/media/1016/draft-guidance-on-submissions-ref-2018_1.pdf

³ <https://unlockingresearch-blog.lib.cam.ac.uk/?p=2115>

The repository must be registered in the Directory of Open Access Repositories (OpenDOAR) or in the process of being registered.

In addition, the following criteria for repositories are required:

- *Automated manuscript ingest facility*
- *Full text stored in XML in JATS standard (or equivalent)*
- *Quality assured metadata in standard interoperable format, including information on the DOI of the original publication, on the version deposited (AAM/VoR), on the open access status and the license of the deposited version. The metadata must fulfil the same quality criteria as Open Access journals and platforms (see above). In particular, metadata must include complete and reliable information on funding provided by cOAlition S funders. OpenAIRE compliance is strongly recommended.*
- *Open API to allow others (including machines) to access the content*
- *QA process to integrate full text with core abstract and indexing services (for example PubMed)*
- *Continuous availability*
- *Helpdesk"*

These requirements, if required by 1 January 2020, will be difficult, if not impossible, to implement especially for smaller institutions already groaning under the demands of other funder policies. In the UK the repository community is relatively well equipped to respond to the challenges of Plan S, yet we have concerns for our international colleagues in institutions yet to deploy or fully embed an institutional repository in their research services.

In addition to the specific concerns addressed in the Plan S feedback questions (see Appendix), we would also note that the majority of gold open access journals do not levy APC charges. These titles are an important venue for communicating research outcomes in specific disciplines and their existence is very much in the spirit of Plan S. Yet, technical barriers have been raised by Plan S such that very few of these titles will satisfy the technical requirements of Plan S, thereby once again excluding some scholars from scholarly discourse.

While we have outlined a number of criticisms of the Plan S implementation guidance for repositories, it is important to reiterate that the overriding aims and ambitions of Plan S are not without merit. Repository managers will be looking for further clarity from cOAlition S on what these requirements actually entail. Certainly the most difficult will be the conversion of manuscripts into full text XML. We believe this should be an aspirational recommendation.

Dr Arthur Smith on behalf of UKCORR

Secretary to UKCORR

Deputy Manager of Scholarly Communication (Open Access)

Office of Scholarly Communication

Cambridge University Library

West Road, Cambridge CB3 9DR

as2507@cam.ac.uk

Appendix

Guidance on the Implementation of Plan S – Feedback Questions

1. Is there anything unclear or are there any issues that have not been addressed by the guidance document?

The guidance in relation to open access repositories is highly technical yet misses significant details. We would therefore ask for more clarity from cOAlition S regarding these requirements. We are specifically concerned about the requirement for full text XML which we see as being highly problematic for the vast majority of institutional repositories. We believe this should be an aspirational recommendation rather than a strict requirement of Plan S.

The repository must be registered in the Directory of Open Access Repositories (OpenDOAR) or in the process of being registered.

This requirement is not too onerous, and most institutional repositories should be able to comply with this requirement.

Automated manuscript ingest facility

cOAlition S should more carefully define this requirement, as it is not immediately obvious what is required. Amongst the many interpretations, this requirement could refer to:

- All manuscripts deposited in repositories must be automatically ingested (i.e. no author self-archiving).
- Ingest of accepted manuscripts from authors via an online form.
- Ingest directly from publisher systems, where content and metadata is supplied directly to repositories.
- Automated conversion of content to suitable storage and preservation formats.

Services such as the Jisc Publications Router⁴ (a UK initiative) are being developed which offer this functionality, however, publisher coverage and institutional engagement is limited. While this may improve with the implementation of Plan S, it is not guaranteed.

Some publishers, such as Biomed Central and Hindawi, already provide SWORD⁵ deposits of papers directly to institutional repositories, though these services can be tied to membership deals which have an associated cost. With few legacy publishers currently represented within initiatives like Jisc Publications Router, some journal titles - notwithstanding any hybrid restrictions - may needlessly reduce the repository compliance routes available to authors, particularly those in the arts and humanities, whose publishers are less likely to offer this service.

⁴ <https://pubrouter.jisc.ac.uk/>

⁵ [https://en.wikipedia.org/wiki/SWORD_\(protocol\)](https://en.wikipedia.org/wiki/SWORD_(protocol))

Full text stored in XML in JATS standard (or equivalent)

Most authors publish manuscripts using Microsoft Word, LaTeX or through a publisher's online manuscript submission system. In the UK, authors typically self-archive their accepted manuscripts in repositories in either Word or PDF format. Conversion of these document formats to structured XML is not straightforward. Europe PMC, for instance, outsources the XML conversion of accepted manuscripts⁶, and the manuscripts that institutional repositories typically receive often lack the requisite structure to generate reliable JATS output. This will often be the result of poorly structured publisher manuscript templates or the product of an author lacking awareness surrounding digital formatting and publishing.

This requirement would mean that all but the very best resourced repositories would fail to meet the requirements of Plan S by 2020. Indeed, this requirement is so difficult to implement that many repositories may simply never meet this requirement without significant investment from Plan S funders. Much more clarity is needed from cOAlition S.

We certainly agree that having full text stored in this way will facilitate text and data mining (TDM), however, for the majority of researchers no one repository or publishing platform holds the total sum of the content that they might wish to use. Therefore the preferred approach to TDM is to use external services to extract and aggregate the content before processing the text and/or data. In this situation it is sufficient to mandate that the full text is available for automated discovery by other platforms which can themselves automate text extraction and analysis.

We strongly urge cOAlition S to reconsider this requirement. At the very least, we believe this should be a recommendation for repositories to aspire to, rather than a strict requirement. For many institutions this requirement will mean their repositories are never Plan S compliant.

Quality assured metadata in standard interoperable format, including information on the DOI of the original publication, on the version deposited (AAM/VoR), on the open access status and the license of the deposited version. The metadata must fulfil the same quality criteria as Open Access journals and platforms (see above). In particular, metadata must include complete and reliable information on funding provided by cOAlition S funders. OpenAIRE compliance is strongly recommended.

Most institutional repositories will meet this requirement. However, some work may be needed to implement the necessary modifications for OpenAIRE compliance.

Open API to allow others (including machines) to access the content

In the UK, most institutions operate repositories using either EPrints⁷ or DSpace⁸ software⁹, although some institutions also use a front-end to their current research information system (CRIS) to substitute for a repository. Examples of this latter configuration tend to use Elsevier's Pure Portal product¹⁰. All of these systems provide OAI-PMH¹¹ interfaces, facilitating a level of machine access to

⁶ <http://blog.europepmc.org/2019/02/europe-pmc-plan-s-feedback.html>

⁷ <https://www.eprints.org/uk/>

⁸ <https://duraspace.org/dspace/>

⁹ <http://irus.mimas.ac.uk/about/participants/>

¹⁰ <https://www.elsevier.com/en-gb/solutions/pure>

¹¹ <https://www.openarchives.org/pmh/>

content which also underpins some of the world's largest open access aggregation systems, including CORE¹² and BASE¹³.

Since version 4.0 DSpace has offered an API which can also be used to query repository content, as does EPrints (since version 3.2). Elsevier's Pure CRIS also provides API access, although as an API for a CRIS the interface supports a wider range of query types for content or data repositories would not store.

As noted, repositories are relatively well placed to respond to this Plan S technical requirement. Indeed, OAI-PMH has been very successful at exposing open repository content and furthering open scholarship. However, the success of OAI-PMH was predicated on standards and interoperability. OAI-PMH is a technical protocol and standard to which a variety of disparate systems can subscribe. The Plan S requirement for an 'open API', on the other hand, is too vague to support high levels of machine level access to repositories. All systems implement their APIs in different ways, and any further extensions to repository APIs are likely to diverge too. UKCORR would therefore welcome Plan S initiatives to find consensus on repository machine interfaces to ensure resources invested into API enhancements are maximised to the benefit of a common open scholarly communications infrastructure and open research more generally.

QA process to integrate full text with core abstract and indexing services (for example PubMed)

This is certainly possible, however, this requirement relies on abstracting and indexing services to actually allow repositories to easily feed content into their systems. So far this has been very limited. To this extent repositories would again be dependent on the technical support of third party services in order to assure their relevance as a compliance route under Plan S.

These guidelines give no indication as to whether these services should be free. If not, institutions may need to pay abstracting services to publish content to them which would surely not be in the spirit of Plan S, since previously open content would then be open to monetisation by the very legacy publishers whose influence Plan S is seeking to curtail.

Further clarity from cOAlition S regarding the specifics of this requirement would be welcomed.

Continuous availability

As a matter of course most institutional repositories endeavour to meet this requirement, however, few would guarantee availability with a service level agreement. Such a step would require significant investment and is probably beyond the scope of most institutions.

Helpdesk

Institutions, through their library or IT offices, usually provide some level of support for repository users, though there is often no guaranteed level of service.

2. Are there other mechanisms or requirements funders should consider to foster full and immediate Open Access of research outputs?

We would strongly urge cOAlition S to place a greater emphasis on green open access as a suitable route to meet the requirements of Plan S. The current guidance favours gold open access above all

¹² <https://core.ac.uk/>

¹³ <https://www.base-search.net/>

else and diminishes the role that open access repositories have played in driving forward the open research agenda. Authors should be able to choose how their research is made open access. Open access repositories add to that choice, and by unnecessarily stifling the ability of open access repositories to serve the research community, Plan S may actually reduce diversity and increase costs.