Plan S Implementation Guidelines for Repositories
Results of COAR Repository Platforms Survey

June 17, 2020

Prepared by Kathleen Shearer, Executive Director, COAR

(for more information contact: office@coar-repositories.org)

Executive Summary

In 2019, a group of funders known as cOAlition S adopted Plan S, a set of principles and requirements for full and immediate Open Access to peer-reviewed scholarly publications resulting from the research they fund, beginning in 2021. One of the routes for complying with Plan S is for authors to make the final published version (Version of Record, VoR) or the Author's Accepted Manuscript (AAM) openly available with an open license in a Plan S compliant repository with immediate OA from the date of publication.

In order to support compliance with Plan S, repository software platforms, repository managers and researchers (who use the repositories) will need to be aware of the requirements and, in some cases, adopt new practices and functionalities. In April/May 2020 the Confederation of Open Access Repositories (COAR), in consultation with cOAlition S, conducted a survey of repository platforms in order to assess their current ability and intention to support Plan S requirements, and to identify any specific challenges related to their implementation.

The survey found that most repository platforms currently support compliance with Plan S mandatory criteria and, in the few cases where they do not, there are plans to adopt this functionality. In addition, many of the highly recommended criteria are also already supported by the platforms. As a next step, COAR and cOAlition S will continue to work together to ensure that repositories are well represented and develop more detailed guidance that assists them in supporting the major functionalities envisioned in Plan S.

Introduction

Repositories are essential services that manage and provide access to articles, data, and many other types of research outputs. They offer a low-cost, high-value option for providing open access, and are also a mechanism for introducing innovation in scholarly communication. With over 5000 repositories around the world, the international repository network represents critical infrastructure for open access and open science.

In order to support compliance with Plan S, repository software platforms, repository managers and researchers (who use the repositories) will need to be aware of the requirements and, in some cases, adopt new practices and functionalities. In April/May 2020 COAR, in consultation with cOAlition S, conducted a survey of repository platforms in order to assess their current ability and intention to support Plan S requirements, and to identify any specific challenges related to their implementation. This survey, developed in conjunction with a cOAlition S Task Group[1], is part of an ongoing collaboration between COAR and cOAlition S to ensure repositories are appropriately represented in Plan S and can adhere to requirements.

Participating repositories

Sixteen repository platforms responded to the survey including most of the major open source platforms, as well as several of the large national, regional and domain repositories. The results of the survey represent a large portion of the open access repositories currently in use around the world for sharing scholarly publications.

Table 1: Repositories / software platforms that responded to the survey

Domain repositories	National repositories	Catch-all repositories	Repository software platforms hosted by more than one institution
- arXiv - INFN Open Access Repository - PubMed Central	- Episcience - HAL - Jisc Open Research Hub	- Figshare - Zenodo	- Brocade - DSpace - EPrints - Fedora - Islandora - LibreCat - OPUS - TIND IR

---

[1]Task Group Members: Jon Øygarden Flæten, The Research Council of Norway; Zoe Ancion, Open Science Project Leader, French National Research Agency; Jyrki Hakapää, Academy of Finland; Sally Rumsey, Jisc/cOAlition S

Survey Results

Plan S Mandatory Criteria for Repositories

> › Use of PIDs for the deposited versions of the publications (with versioning, for example in case of revisions), such as DOI (preferable), URN, or Handle.
>
> › High quality article level metadata in standard interoperable non-proprietary format, under a CC0 public domain dedication. This must include information on the DOI (or other PIDs) both of the original publication and the deposited version, on the version deposited (AAM/VoR), and on the Open Access status and the license of the deposited version. Metadata must include complete and reliable information on funding provided by cOAlition S funders (including as a minimum the name of the funder and the grant number/identifier).
>
> › Machine readable information on the Open Access status and the license embedded in the article, in standard non-proprietary format.
>
> › Continuous availability (uptime at least 99.7%, not taking into account scheduled downtime for maintenance or upgrades).
>
> › Helpdesk: as a minimum an email address (functional mailbox) has to be provided; a response time of no more than one business day must be ensured.

The requirements of Plan S aim to ensure that repositories support widespread discovery, access and reuse of articles by humans and machines as well as monitoring of open access compliance by funders.

The survey asked about the ability of platforms to support the first three mandatory requirements of Plan S (the other two requirements would be the responsibility of the supporting organization, rather than the software platform, they were not included in the survey).

## Graph 1: Responses to mandatory requirements questions
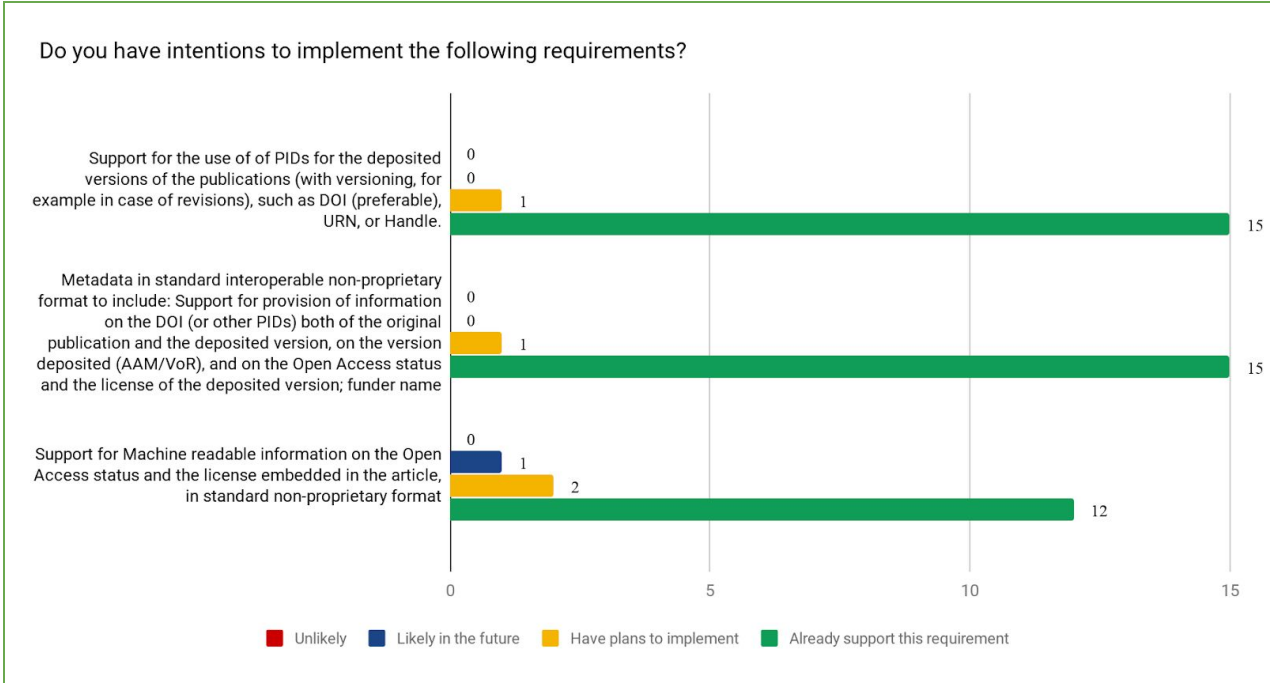
Plan S strongly recommended additional criteria for repositories

> Manuscript submission system that supports both individual author uploads and bulk uploads of manuscripts (AAM or VoR) by publishers.

> Full text stored in a machine-readable community standard format such as JATS XML.

> Support for PIDs for authors (e.g., ORCID), funders, funding programmes and grants, institutions, and other relevant entities.

> Openly accessible data on citations according to the standards by the Initiative for Open Citations (I4OC).

> Open API to allow others (including machines) to access the content. A compliant API must be free to access without any barrier. A light authentication mechanism such as a token for 'power users' – e.g., high-traffic collaborators – is acceptable as long as there is a totally open/anonymous route too.

> OpenAIRE compliance of the metadata.

> Quality assurance processes to link full-text deposits with authoritative bibliographic metadata from third party systems, e.g., PubMed, Crossref, or SCOPUS where feasible.

Many of the strongly recommended criteria are also supported by the repository platforms, with the exception of a couple, which ask for functionalities that are not common practice in repositories: JATS XML and Open Citations.

Graph 2: Responses to strongly recommended criteria questions

Do you have intentions to implement the following recommendations?
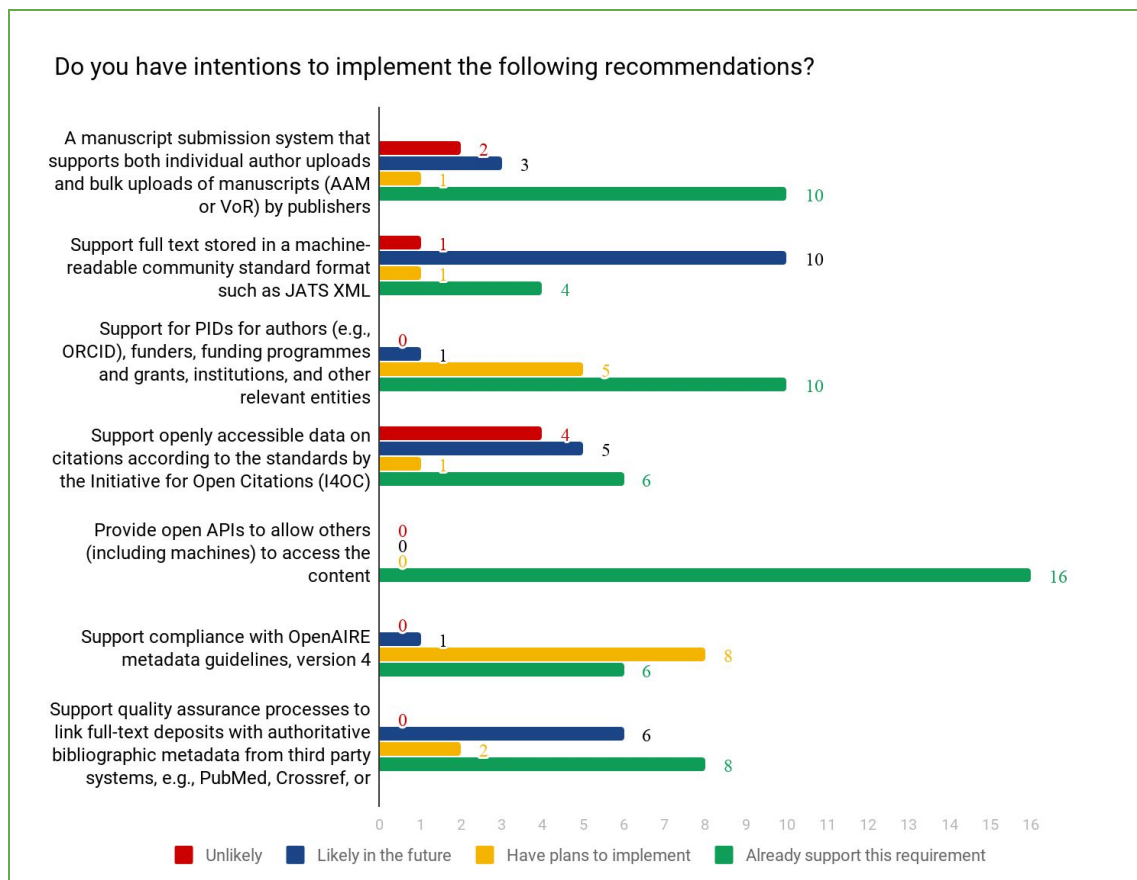
COAR Findings and Recommendations

The results of the survey show that most repository platforms already support compliance with Plan S mandatory criteria and in the few cases, where the platform does not, there are plans to adopt the functionality. Several respondents noted that the mandatory requirements are quite vague and could be interpreted differently across repositories. It would be helpful, therefore, to provide the community with more detailed guidance about the mandatory requirements, particularly in terms of "high quality metadata" and "interoperable formats".

COAR also advocates for Interoperability and quality metadata in repositories, because those practices support greater discovery and reuse of articles. Currently, repositories around the world are subject to a variety of requirements related to metadata including regional, national and domain requirements (FAIR, OpenAIRE, RIOXX, JPCOAR, and so on). However, currently many repositories support only basic metadata elements that are the default in the repository software. It will take some effort for individual repositories to adopt more comprehensive and granular metadata elements, such as DOIs for funder, institution, as well as open access status. COAR has been working with the repository community in order to improve metadata in repositories, and will be launching a campaign to raise awareness of the benefits of good quality metadata. In addition, the COAR has developed some basic recommended metadata elements which can be adopted in a variety of schemas in order to support interoperability.

In terms of the requirement to adopt machine readable licenses and open access status embedded in the article, while most repositories do support this, this is not something that would be typically found in an Accepted Manuscript or Version of Record. In order to support machine access to the full text of articles, [COAR](#) recommends that repositories adopt the [Signposting protocol](#) as the standard convention for making full text easily discoverable and retrievable by external services and processes. Signposting defines a common approach for including information about the Author, Bibliographic Metadata, Identifier, Publication Boundary, and Resource Type in the landing page of the repository (or other type of data provider).

While the Plan S mandatory criteria are well supported by the repository platforms already, there are lower levels of support for the highly recommended criteria in the repository platforms.

A manuscript submission system that supports both individual author uploads and bulk uploads of manuscripts (AAM or VoR) by publishers

Although all repository platforms support individual uploads (commonly referred to as deposits, in the repository community), and many do already offer the functionality of bulk uploads, there are some platforms that do not support this, and do not plan to in the future. This functionality has been a point of debate in the repository community, as some feel that repositories should not be dependent on publishers for content provision, while others feel this could be a pragmatic and effective approach to populating repositories. Moreover, many publishers will be willing to

deposit (bulk upload) into large domain repositories such as PubMed Central, but arre unlikely to provide this same functionality for institutional repositories. Intermediary services such as as brokers/routers (e.g. Jisc Publication Router), which gather information from content providers such as publishers, passing it on to institutions to help them capture their research articles onto their systems, could be a solution for repositories to capture a larger number of articles, without relying on individual authors.

Support full text stored in a machine-readable community standard format such as JATS XML

Full text stored in a machine-readable community standard format such as JATS XML, which enables text and data mining (TDM) of articles, is not commonly supported in repositories at the moment, with exception of a few larger repositories, such as PubMed Central and Europe PMC. In most repositories articles are either in pdf or word formats. Converting articles to xml from word/pdf is extremely resource intensive and it is unlikely that this functionality will be adopted in many repository platforms. TDM, however, is generally supported by repositories because they allow their full text articles to be harvested and indexed by large scale networks (e.g. CORE, OpenAIRE), which already perform text and data mining on their aggregations.

Support for PIDs for authors (e.g., ORCID), funders, funding programmes and grants, institutions, and other relevant entities

Most repository platforms currently support the adoption of PIDs or plan to do so in the future. In particular, all platforms support inclusion of ORCID in their metadata records, but many do not yet support other PIDs such as funder, funding program, and institution. This issue is related to the default metadata schema included in many platforms, which does not include these fields. For repositories that do not currently support this type of metadata, an alternative approach using an alternative metadata field (such as in the "sponsorship" or "contributor" field in other platforms) can be recommended so common practices are adopted by the different repositories around the world.

Support openly accessible data on citations according to the standards by the Initiative for Open Citations (I4OC)

Six of the fifteen responding repository platforms currently support openly accessible data on citations. And while this is a highly desirable functionality as it will allow the community to develop non-proprietary citation tools and enable aggregation of citations across article versions, repositories will have difficulty complying with the specific recommendations of the Initiative for Open Citations (I4OC) unless the version of the article being deposited has already been formatted appropriately, and the resources required to process citations in each article may be too high for many repositories. There are, however, other ways of supporting open citations in the context of repositories, through the use of network services. Zenodo, for example, enables linking records with related persistent identifiers (both incoming and outgoing links - i.e. citations

and references), and these links are exported in the DataCite metadata registered with the DOI, and then available in the CrossRef/DataCite Event Data. These types of solutions can be expanded to other repositories to support the goals of open citation, but will unlikely be fully implemented by most platforms by 2021.

Provide open APIs to allow others (including machines) to access the content

Open APIs allows harvesting and indexing of metadata and full text content in repositories. An OAI-PMH feed is standard for most repositories and all major repository platforms and all respondents indicated that they currently support Open APIs.

Support compliance with OpenAIRE metadata guidelines, version 4

Many platforms are working towards compliance with OpenAIRE metadata guidelines for literature repositories. Because version 4 of the guidelines will require some extra effort and development on the part of the platforms (because they require the adoption of entities or subproperties of metadata for funder name, funding stream and award number), it may take some time to support the adoption of version 4. In addition, because OpenAIRE guidelines were developed to support tracking of European Commission funding information, it may not be a priority for some repository platforms that have an international user base. As an alternative, some more simple metadata can be recommended that will enable tracking of open access content and ensure interoperability and discovery.

Support quality assurance processes to link full-text deposits with authoritative bibliographic metadata from third party systems

Indexing of repository content through third party sources is important for discovery and tracking. While most repositories do not contribute their metadata to Crossref, they are regularly indexed / harvested by national and regional aggregators, such as OpenAIRE, CORE, BASE, LA Referencia and so on. These national and regional harvesters, exchange data with each other, and therefore are the best indexes of repository content.

Next steps

The survey results indicate that most of the highly used repository platforms already support Plan S mandatory requirements, as well as many of the strongly recommended criteria. There are some areas in which they could use further support, clarification or more detailed instructions. In addition, it's important to note that even if the repository platform supports Plan S requirements, there will be additional efforts required at the level of the local repository and also by the author/researcher depositing the article in order to comply with Plan S.

To ensure that repositories are able to broadly support the major functionalities envisioned by cOAlition S (machine and human reuse, text and data mining, and tracking of OA content), COAR and cOALition S will continue to work together to develop more detailed guidance in several areas that will be helpful for repository platforms, as well as researchers and individual repository managers. Some of these instructions will be made available shortly on cOAlition S website. In addition, we will examine some of the more challenging recommendations (JATS-XML and open citations) and assess whether an alternative approach to these functionalities may be more appropriate for repositories.

About COAR

The Confederation of Open Access Repositories (COAR) is an international association with members and partners from around the world representing libraries, universities, research institutions, government funders and others. COAR brings together individual repositories and repository networks in order to build capacity, align policies and practices, and act as a global voice for the repository community.

About cOAlition S

cOAlition S funders (a group of national research funders, European and international organisations and charitable foundations) have agreed to implement the 10 principles of Plan S in a coordinated way, together with the European Commission and the ERC. Other research funders from across the world, both public and private, are invited to join cOAlition S