

Principles for establishing a Central Service for Preprints: a statement from a consortium of funders

Preprints: a definition

Complete and public drafts of scientific documents, yet to be certified by peer review.

Preamble

At the ASAPbio¹ Funders [Workshop](#) (held on the 24th May 2016) broad agreement was reached on the value of a Central Service (CS) for preprints. Although the detail of the proposed CS was not fully determined, there was support for the view that a future CS would **aggregate** content from multiple sources² and provide new ways for researchers and machines to **search**, **access** and **reuse** this content. The CS would also have an **archival** function, ensuring long term, stable access to preprints. To help realise these different functions - archival, access, re-use etc - there is assumption the CS would also **convert** all ingested preprints to a standard file format.

The Funders listed below believe that sharing of preprints provides researchers with a faster way to disseminate their work, establish priority of their discoveries, acknowledge funders' contributions to research advancement, and obtain feedback. They also offer a more current understanding of an investigator's work.

We also believe the development of the CS will provide the research community with a crucial resource that will ensure that preprints, regardless of origin and format, can be discovered, accessed and used.

As a consequence of these factors the Funders are highly supportive of the work ASAPbio has been doing to encourage preprint sharing in the life and biomedical sciences and, more recently, their work to start to define the key elements of the proposed CS.

While the Funders (listed at the end of this document) are not committing themselves to fund a CS, we strongly encourage ASAPbio to develop a proposal describing the governance, infrastructure and standards desired for a CS that represents the views and needs of the research community, **which includes both researchers and funders**. The proposal should include a budget, goals, milestones, implementation timeline and sustainability plan after 5 years of funding to bring an appropriate community-defined preprint CS into a stable, long-term service.

¹ [ASAPbio](#) is a scientist-driven initiative to promote the productive use of preprints in the life sciences

² This would include existing preprint services, such as [bioRxiv](#) and [PeerJ Preprints](#), and future services, perhaps those established by publishers or other discipline-based services (e.g. [ChemRxiv](#)).

To help frame the proposal - and understand the intent of the Funders with regard to establishing a CS - the Funders have drawn up this document which articulates a number of **principles** and, in some cases, **requirements**, which will need to be adhered to as a condition of any future funding.

Developing a Central Service for Preprints: overarching principles

Principle 1: The Central Service must have an independent governance structure

1.1 Governance: overview

We support the notion that the CS should be governed by an independent governance body that is international in scope and led by highly respected members of the research community, and includes other relevant experts including organizations that serve the research community; policy and legal experts; and technical experts. For the purpose of this document we will assume that ASAPbio is the entity which is responsible for managing the CS. As a consequence, we also assume that the primary decision making body will be an ASAPbio Board of Directors.

This body will be responsible for defining the details of the service - what is an acceptable repository from which content can be aggregated and what types of preprints are within scope - and determining a longer term sustainability model for the CS. We also assume that this body will be responsible for managing and running a procurement process to identify a supplier (or a consortium of suppliers) to deliver this CS.

At this stage it is unclear whether ASAPbio will be the entity which will actually contract/grant fund the supplier(s) of the CS – or whether this is managed by funders of the CS, either collectively or through a designated lead funder. In the event that funding for the CS is managed directly by those funders supporting the cost of the CS, then the ASAPbio Board will be expected to advise the funders on which supplier (or consortium of suppliers) should be funded to deliver the CS.

To help ensure that the services continue to meet the needs of the research community, we envisage that the ASAPbio Board of Directors will be supported by a Scientific Advisory Board (SAB) appointed by, and reporting to, the ASAPbio Board of Directors.

Funder requirement:

The CS must have an independent governance structure.

1.2 Governance: Funder role

The Funders do not envisage having any formal role on the ASAPbio Board of Directors or the SAB; indeed it is critical that the major decision making bodies should be independent of the Funders.

Working on the assumption that a consortium of funders will fund the CS, a mechanism will be needed to ensure that the service being developed is in line with the requirements of the research community.

One mechanism to explore might involve the ASAPbio Board of Directors providing an annual report to the Funders of the CS - outlining what they have done over the past 12 months and what developments are planned for the next 12 months. This report would be developed with input from the Scientific Advisory Board. Funders would use the report as a mechanism for determining whether to release the next 12 months of funding.

Principle 2: The Central Service should seek to secure widespread community support

2.1 Community support

It is essential that ASAPbio engages broadly with the research community to ensure that the CS enjoys as much community support as possible. The Funders will expect ASAPbio and the SAB to continue to engage with the research community to seek their input on the future direction of the CS and to promote its use.

Principle 3: Content in the Central Service should be open and meet scholarly standards

3.1 Preprints should be made maximally useful through permissive licensing

We believe that to maximise the benefits which arise through the sharing of preprints, content made available through the CS should be licensed in ways which facilitates re-use, text and data mining and the development of services which allows others to innovate on this content. As funders we strongly believe that this can best be facilitated by ensuring that the content made available through the CS is licensed using the Creative Commons Attribution licence, CC-BY.

However, we also recognise that if we limit the CS to *only* aggregate CC-BY content, this may adversely impact the uptake of preprints and the CS's intent to be the premier discovery system for preprints.

In the longer term the Funders would like to get a position whereby the CS *only* aggregates CC-BY licensed content. We will work with the CS, and the community more broadly, through ASAPbio, to determine the most effective policy levers to bring this about.

Funder requirements:

All preprints made available through the CS must include a licence statement, which makes it clear how that content can be used.

All aggregated content must be included in the full text corpus for search (e.g. showing snippets like Google books) and be made available for text and data mining and other computational uses (via the CS API).

Content providers - including existing preprint servers, publishers and users who post directly to the CS (if that is deemed to be a useful service) - who want their content to be discoverable through the CS - must agree to these conditions.

3.2 Preprint - the underlying data

As Funders we recognise the importance of making the underlying data - referenced in a preprint - available for others. Sharing data reduces waste, supports reproducibility and helps accelerate discovery and its application for health benefit.

Consequently, we strongly encourage both existing and emerging preprint servers to develop their services such that, going forward, all preprints (which are aggregated by the CS) include a data availability statement. We also encourage researchers to make the underlying data available under a CC-BY or the Creative Commons [Public Domain Dedication waiver](#) (CC Zero), at the time of formal (peer reviewed) publication, provided that this is consistent with any commercial, legal and ethical obligations.

In terms of data availability we strongly encourage researchers to deposit data in recognised repositories. Where these do not exist, we encourage researchers to make their data available via more generic repositories, such as [Dryad](#), [Figshare](#) and [Zenodo](#).

3.3 Preprint - scholarly standards

As funders we feel that it is important that the CS must uphold scholarly standards of publication. Preprints ingested into the CS must adhere to standard scholarly publication practices such as authorship, regulation and ethical, legal and societal standards. They must also provide appropriate funding acknowledgements. In addition, the CS workflow must support mechanisms (e.g. screening of content) to ensure that these standards are maintained.

Funder requirement:

The ASAPbio Board must develop a clear set of guidelines to ensure that content aggregated into the CS upholds scholarly standards of publication.

Principle 4: Where possible, the CS should make use of and build on existing infrastructure, services and good practice

4.1 Build on existing infrastructure and services

A number of relevant services, tools, and applications already exist which potentially could be used to support the development of the CS. Where appropriate, we encourage ASAPbio to issue a proposal that fosters relations with these providers of services, tools, and/or applications, so as to maximise support and collaboration from existing “preprint communities”.

In terms of best practice, we would expect the CS to make use of proven technology (one example might be the use of [JATS XML](#) for document conversion) but at the same time keep an open mind to experiment as new opportunities emerge.

Principle 5: Any new code to build the Central Service should be open and interoperable

5.1 Central Service: software

As Funders we wish to create a vibrant preprint ecosystem to help advance the use of scientific publications. We believe that we will best achieve this by adopting an open source licensing model.

Funder requirement:

Any software which is used or developed to support the CS should be made available under open licenses, such as those developed by MIT or BSD. If, downstream, a supplier responding to any Request for Information (RFI) or tender request is not able to comply with this approach they will need to explain why, and what public benefits will be realised by adopting a less open licensing regime.

5.2 Central Service software should support and foster interoperability

As Funders we believe that the CS will only be successful - that is making a critical mass of preprints available in ways which allows others to build and innovate on this content - if any system that is developed is built with interoperability as a key guiding principle.

Specifically, we believe that the CS will need to interoperate with other systems - ingest servers, screening services, metadata and utilization statistics, etc - and that a system that does not support open APIs is unlikely to succeed. And, though the CS should be limited in scope to preprints in the life and biological sciences, we should be mindful that as research becomes more interdisciplinary it may be desirable to bring in preprints from other disciplines to create an “all scholarship preprint service”, or, at the very least, allow others to use the software developed by the CS to establish their own services.

Principle 6: Access to the Central Service must be free at the point of use

6.1 Free, unfettered access

To foster uptake of preprints amongst the life sciences research community we believe that access to the CS must be free at the point of use for both suppliers and consumers of content.

Funder requirement:

Access to the CS must be free at the point of use for both suppliers and consumers of content.

Principle 7: the Central Service must be easy to use

7.1 Easy and rewarding to use

For the CS to be successful it must be easy for researchers, developers, publishers etc. to use and engage with. By way of example, it must be possible for the CS to aggregate content directly - either from existing preprint servers or from publishers who wish to make submitted manuscripts available

to the CS. If it deemed important for the CS to offer a “direct deposit mechanism” (so researchers can post directly to the CS) then this must also be easy to use.

Equally, the CS should provide a rich search and discovery experience so that researchers can identify, access and, where available, download content that is most relevant. Finally, it should offer usage/impact metrics and facilitate incentives that reward researchers for posting preprints and other service activity (commenting, data sharing, screening, etc.).

Beyond these core features, we encourage ASAPbio to work with the community to help better understand how researchers might want to use the CS and what services need to be developed to support these needs. We believe this represents an unprecedented opportunity to further scholarly communication in the life and biomedical sciences.

7.2 Easy for developers and other applications to use

In addition to researchers, we believe that key users of the CS will be other developers and applications who wish to build rich, value-added services on top of the CS. To enable this, the CS would need to develop a suite of open APIs to support these capabilities.

Funder requirement:

The CS must provide open APIs to allow others to build new services based on the content it has aggregated.

Principle 8: The Central Service must have a sustainable model

8.1 Sustainability

The proposal to develop the CS should include a credible plan to develop and sustain an appropriate community-defined CS for preprints.

Funder requirement:

A credible sustainability plan is required, to demonstrate how the CS will support itself in the long term.

8.2 Cost effective and flexible approach

The research community, including funders, are already spending significant sums of money on scholarly communication (mainly in the form of subscriptions and open access costs). Given this, the CS proposal would represent another cost for Funders, at least in the short term. Consequently, it is essential that the approach taken be cost effective and flexible (so that it can adapt to changes in the ecosystem and stakeholder needs) and that any aspect of the CS that is not adding value is eliminated.

Funders supporting these principles

Funder name	Funder representative
Wellcome Trust	Robert Kiley
National Institutes of Health	Patricia Flatley Brennan
Medical Research Council (UK)	Tony Peatfield
Helmsley Trust	Megan Deichler
Howard Hughes Medical Institute (HHMI)	Judith Glaven
European Research Council	Dagmar Meyer
Simons Foundation	John Spiro
Canadian Institutes for Health Research	Matthew Garsia
Alfred P. Sloan Foundation	Josh Greenberg
Department of Biotechnology, Government of India	K. VijayRaghavan
Laura and John Arnold Foundation	Mike Stebbins

This document was prepared by Robert Kiley, Wellcome and Philip Bourne, NIH (Associate Director of NIH for Data Science, 3.14-1.17), with additional input and comment from Stuart Buck (Arnold Foundation), Lorraine Egan (Damon Runyon Cancer Research Foundation) Sindy Escobar-Alvarez (Doris Duke Charitable Foundation), Michele Garfinkel (EMBO), Josh Greenberg (Sloan Foundation), Maria Leptin (EMBO), Tony Peatfield (Medical Research Council) John Spiro (Simons Foundation), and Neil Thakur (NIH).