Request for Applications

Central Service for Preprints in the Life Sciences

Table of Contents

This is a Request for Applications for the development of a Central Service (provisional name) for preprints in the life sciences issued by ASAPbio. This Request is open to all prospective bidders, and we encourage responses from interested parties able to deliver the services described below. For a concise description of the goals of this project, please see our blog post entitled The Benefits of a "Central Service" for Biology Preprints.

1. ASAPbio's Overall Goals Relevant to this RFA

1A. Creating a Central Service that serves the public good

Capitalizing on the interests of many of the world's leading funding agencies in the life sciences (see Appendix 1. Key Principles for the Central Service developed by the Funders' Consortium), we are poised at a moment of incredible opportunity to create next generation infrastructure for the exchange of scientific information in the form of preprints and related pre-peer reviewed content. The Central Service (CS) will be a database of the full text, associated metadata, figures, tables, and supplementary materials of preprints in the life sciences. This material will be available in single corpus that is open to both human and machine access, which will facilitate the development of 3rd party applications and services. Funding agencies and universities also could look to the Central Service as a single vetted source of preprint material for grant applications and tenure promotions. The creation of a Central Service also provides an opportunity for the scientific community, in the form of a Governance Body, to oversee the standards and developments of preprints over time.

ASAPbio views preprints as public goods that enhance and accelerate scientific progress. In an effort to determine how preprints could best serve the scientific community and advance science as a whole, ASAPbio has conducted a series of meetings with numerous stakeholders in the scientific enterprise (Appendix 2. Introduction and Background Leading to this RFA). Through these discussions, consensus has developed that a Central Service for preprints governed by the scientific community would be timely and would advance scholarly communication in the life sciences. While new innovations and business models utilizing the open data contained within the Central Service might emerge, the purpose of the Central Service is neither to benefit nor to enhance the stature or profile of any particular organization. Rather, the role of the Central Service is to serve the scientific community and the public that funds biological and life sciences research. Through this RFA, ASAPbio is looking for service providers who are both experts in their work and aligned with the mission and vision of ASAPbio. The service provider(s) also must be willing to work with the scientific community in the form of an elected Governance Body (GB) that will oversee the work of the Central Service. Since we cannot anticipate all of the work and modifications of the Central Service that might be needed over time, the service provider(s) must be willing to cede autonomy and be willing to work with the scientific community in a two-way dialog that promotes the development of the database and its interactions with complex scientific and publishing enterprises.

Successful responses to this RFA must embrace a spirit of openness and transparency. The funders have articulated a key message that the software used by the Central Service must be open and interoperable, with any exceptions appropriately justified. All of the data and appropriately anonymized usage metrics of the Central Service must be easily available to others through open APIs. A desire for CC-BY licensing for all preprints harvested by the Central Service was expressed, but ASAPbio appreciates that this issue is complex and also requires acceptance by scientists, which must develop over time. The reports of the Central Service and communications from the Governance Body (GB) and the funders will also be made public on the ASAPbio website.

We see the opportunity for this RFA to spark collaborations leading to consortia of providers that work together in the development of the Central Service. The funders state a desire to "build on existing infrastructure, services and best practices." A single supplier might be able to fulfill the Central Service requirements described below. However, we encourage collaborations where strengths and weaknesses of individual organizations could be complemented and result in enhanced functionality of the CS and a reduction in the time to execution. Details of collaborations and associated management structure should be given in the application. If you are looking for collaborations with other organizations for the purpose of completing a proposal, please leave a message on the public ASAPbio Central Service bulletin board google doc.

We anticipate five years of funding for the service, renewable on a yearly basis. The expectation is that the service should be free of charge to all parties (including reading, depositing, and accessing data) for five years contingent on funding availability and guidelines provided by the Governance Body. Beyond five years, we expect ASAPbio, the GB, and service providers to devise a long-term sustainable business model.

1B. Creating and defining the roles of a Governance Body, Funder Consortium, and Central Service Provider(s)

In this vision for creating a Central Service, several parties will need to work together in a cooperative manner to develop, maintain, and evolve the CS. We define these entities and their roles below.

ASAPbio:

- Is a non-profit corporation whose mission is to advocate for the productive use of preprints in biology and to administer grants/contracts for the operation of the Central Service
- Collects and disburses funds to CS provider(s) according to instructions from funders and the Governance Body (GB)
- Acts as secretariat to the GB
    - Organizes GB meetings
    - Provides administrative support in drafting GB reports and recommendations
    - Communicates with the public and other stakeholders to report and seek feedback on GB policies
- Advocates for best practices regarding preprints in general (including their use in grant applications, licensing, journal policies, etc.)
    - These activities are overseen by the ASAPbio Board of Directors, which is separate from the GB
- Prepares an annual report containing a summary of its own activities for funders
- Assists with the preparation of the GB report and its presentation to funders

The Governance Body (GB):

- Is an independent body supported by funds and administrative services provided by ASAPbio. We are anticipating that the GB will have bylaws established and begin operations in September 2017.
- Makes determination on the features and functionality of the Central Service
- Sets standards for the operation of the Central Service
    - A list of eligible ingestion sources
    - Required and optional metadata
    - Screening requirements and protocols
    - Licensing options and requirements

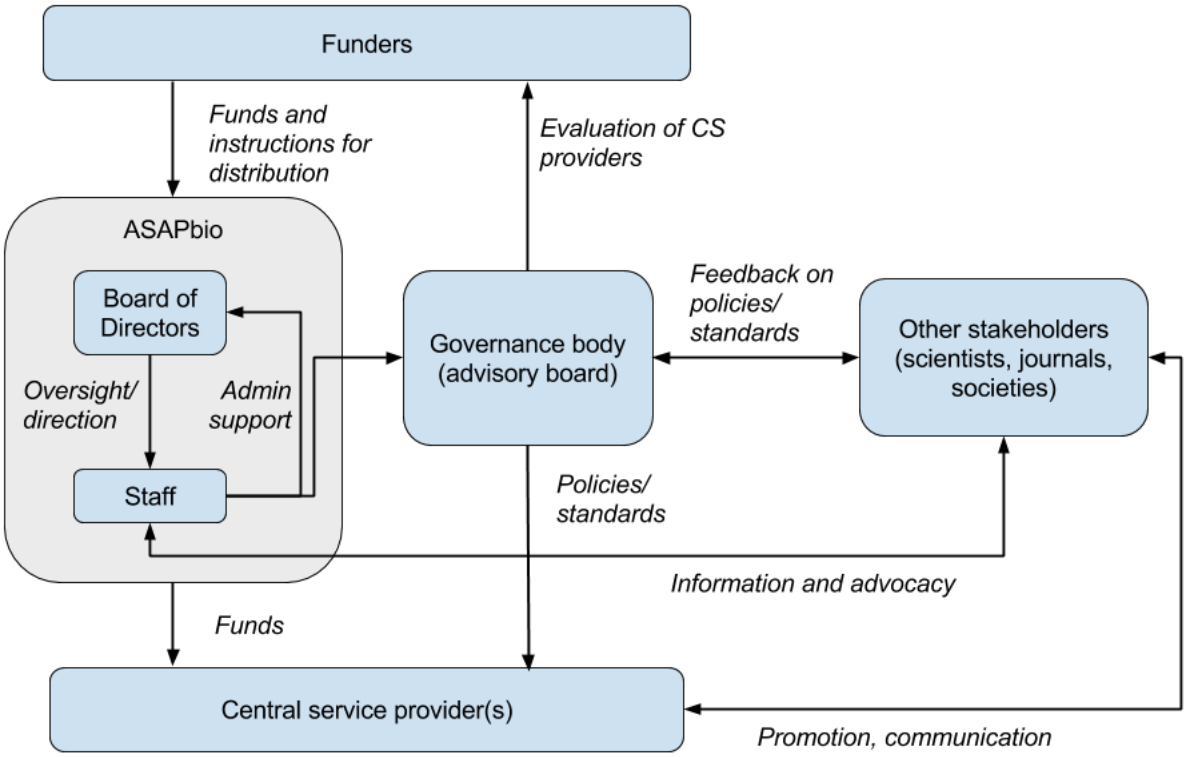
    - Scope of content
    - API protocols
    - Usability criteria
    - Recommendations for direct ingestion and display
- Prepares an annual report for funders (coordinated by ASAPbio) containing:
    - A summary of its own activities
    - An evaluation of CS provider performance
    - Recommended upcoming goals and strategies for the CS
- To perform these functions, the GB seeks input by various means:
    - Directs ASAPbio to run surveys, consultations, or meetings for stakeholder groups or the public
    - Can elect to form ad hoc committees to advise on technical, legal, ethical, or other issues
        - Members of these committees are appointed by the GB
    - Commissions external evaluations of the ingestion sources for the CS in order to ensure that quality standards are met

The Central Service Provider(s):

- Is/are entities that receive grants or contracts from ASAPbio to provide the implementation of the Central Service
- Promotes its service(s) through advertising, outreach, and communication
- Prepares an annual report for the GB containing:
    - Summary of its activities
    - Recommended upcoming goals and strategies for the CS

The Funder Consortium:

The details as to how the Funder Consortium will operate are still to be determined. As a minimum, we anticipate that the Funder Consortium will provide funding to support the development of the Central Service and release funds annually (for an agreed period of time), taking into consideration the recommendations of the GB.


Funders
Funds and instructions for distribution
Evaluation of CS providers
ASAPbio
Board of Directors
Oversight/ direction
Admin support
Staff
Governance body (advisory board)
Feedback on policies/ standards
Other stakeholders (scientists, journals, societies)
Policies/ standards
Information and advocacy
Funds
Central service provider(s)
Promotion, communication

2. Requirements for the Central Service

2A. Overview of Features

In general, preprints in the life sciences currently exist as author-supplied PDFs. While PDFs are simple to upload and handle, our Technical Workshop, as well as numerous discussions with funding agencies, users, and data management experts, have indicated that PDFs are limiting and that preprints should move towards an XML format. With XML, readers can view text, figures, videos, and other multimedia in a more accessible form on computers and mobile devices, link to references, in addition to being able to download content in well-formatted PDF. Importantly, XML provides opportunities for third parties to create tools to enhance knowledge discovery for scientists, which is a forward-looking goal of the Central Service. The generation of the XML content is not intended to encroach upon journals, which provide valuable services in peer review, content aggregation, and editing, but rather to modernize the accessibility of knowledge provided through preprints. Thus, the creation of open source document conversion tools will be an important component of this project with value to the scientific community as well as intake sources (eg. preprint servers). We recognize the document conversation will require the development of new software tools, which will require time to create and is not expected to be operational at the start of the Central Service (see timeline below).

Another important feature of the Central Service will be the creation of open APIs that will facilitate 1) intake of content from multiple sources and 2) access to this content by any party who wants to provide new services derived from this database to scientists. These services could include tools for discovery, commentary, ranking, etc. A single corpus of machine-accessible content will be simple to use (as compared to collecting material that may be inconsistently maintained, formatted, or exposed from multiple sources). Aggregation and the promotion of innovation are two critical catalytic activities of the Central Service.

We also expect innovation in terms of manuscript screening, as an automated tool could identify the subset of preprints that might require further human screening. Such an open source tool could be beneficial potentially for multiple preprint servers, journals, or institutional repositories across different disciplines.

The Central Service will **NOT** provide peer review, as will be stated in its Operating Principles (to be released), and which clearly demarcates its role from the important services of journals. The Central Service will be a “neutral database” and will not execute peer review or be involved in ranking or rating its submissions.

Two areas about which we receive questions are whether the Central Service will receive direct intake from scientists and whether it will display content in its entirety, will link back to the ingestion source for display, or both. There has been considerable discussion on these issues but there is no complete consensus at this time. Factors that influence these decisions also may change over the coming months or year. Therefore, ASAPbio will defer these decisions to the Governance Body that will be created in the summer and begin operation in the fall. The Governance Body will operate according to its mandate and bylaws to best serve the interests of the scientific community. For now, we request that applicants to this RFA provide provisions for intake and display in their proposal, according to the guidelines described below.

Furthermore, we do not intend for the requirements set out below to be overly prescriptive, and we welcome creative and thoughtful solutions from the bidder(s).

2B. Required components

The CS Provider(s) will be required to support a variety of services, each meeting performance targets outlined in Contract. **All components must be distributed under an OSI-approved license.**

1. CS database

The purpose of this component is to create a comprehensive repository of preprints in the life sciences.

In terms of specific requirements, the CS database must, as a minimum, support the following functionality:

a. Act as a repository for life sciences preprints, which includes:
   i. author's original manuscript (.doc, .tex etc.) and the converted manuscript (XML);
   ii. all files associated with that manuscript, such as figures and any supplementary data; including video and datasets, or links to data stored in appropriate external repositories;
   iii. appropriate metadata (to be defined by the GB, but likely to include author information, PIDs such as ORCID and funder IDs, DOIs for manuscript and associated files, ethical statements, screening history, license information).
b. Includes a stable, long-term preservation strategy.
c. At inception of the CS, the database should be populated with legacy preprints (in whatever format they are held: PDF, HTML, XML etc) and associated metadata from existing approved servers. Conversion of these legacy files to XML could be considered but is not a specific requirement for this application.
d. The provider(s) of these services are also required to remove or flag any manuscripts that violate the standards set forth by the Governance Body (ie, plagiarism, privacy of human subjects, dual use research, etc).

2. CS database: human interface

The purpose of these interfaces are to provide PubMed-like search and discovery tools to human users for manuscripts that have passed screening at ingest sources. The question of whether the CS will display the full text of all manuscripts will be determined by the GB. However, the CS will need to have display capability regardless (for example, to be used in cases where the original source of the manuscript is no longer available).

In terms of specific requirements, the CS database human interface must, as a minimum, support the following functionality. We encourage respondents to this RFA to outline how they would make this site the go-to destination for preprints in the life sciences.

a. Provide a web interface for browsing and searching. Searching should be simple and intuitive but also support advanced searches (by ORCID, FundRef, PDB ID etc).
   i. Display abstracts and links back to source, and potentially, to a version held in the CS database
   ii. Potentially provide download functionality for content held in the CS in a variety of formats: PDF, XML, HTML, etc
   iii. Display snippets (like Google Books) to place full-text search results in context

iv. If the source is not available elsewhere, or with consent of the ingestion source, display the full manuscript and figures/supplementary files. Further policies on display of manuscripts at the CS will be determined by the Governance Body
v. Display a clear statement indicating that the material is a preprint (ie. not peer-reviewed)
vi. Link to other versions of the manuscript elsewhere, especially journal versions, using Crossref metadata or information from other sources
vii. Use schema.org tags
viii. Make available metrics and anonymized usage data to humans.
ix. Be developed in line with good user-experience web principles, be fully responsive (and work across all mobile devices)
x. Support login functionality - via ORCID - such that users could save their searches and other preferences

b. Provide an alerting tool that delivers emails/other notifications to authors when content of interest (by keyword, author, citing their article, etc) appears.
c. Respondents are also invited to highlight other functionality they would suggest the site should support, although all features and functionality of the CS will require Governance Body approval.

3. CS database: machine interfaces

These components are intended to make preprints in CS database easy to transmit to third parties for innovative reuse. In Part A, the API enables any entity to access the full text and associated files of any preprint that has passed screening. In Part B, the screening training set is intended solely to enable the development of third party screening tools.

In terms of specific requirements, the CS database machine interface must, as a minimum, support the following functionality.

a. API for manuscripts
i. Make available the full text of preprints (converted version and author's original files) by RESTful API and bulk download
ii. OAI-PMH endpoint, or equivalent if standards change over time
iii. Capture metrics on API use

b. Screening training set
i. Make available all manuscripts (both those that passed screening and those that did not) and their screening histories for bulk download, with any sensitive information redacted. This corpus will be used to train automated screening algorithms, as described above.

4. Manuscript conversion tool

Our ambition for the CS is to create a database of preprints in structured XML rather than the current standard of PDF. The purpose of this component therefore is to develop a tool that provides reliable conversion of author manuscripts to XML. Recognising that any conversion will need to be proofed and approved by the authors, this tool must support this feedback loop. Ingestion sources (ie existing preprint servers or journals) are welcome to provide their own solution to the challenge of providing manuscripts in this format; the CS-provided conversion tool is intended to serve as a resource to the

community. In your application, please specify reasonable targets for the % of XML-converted manuscripts (from formats you specify) that will not require manual correction.

In terms of specific requirements, the manuscript conversion tool must, as a minimum, support the following functionality:

a. Convert the full text of manuscripts, destined for the CS database, to XML or equivalent), tagged according to JATS4R standards. As a minimum it should be possible to convert a .docx or .tex file to XML. The feasibility of converting a .pdf file to XML should be discussed.
b. Allow authors to preview, proof, and modify the converted XML through an author-friendly interface
c. The manuscript conversion tool must be able to interoperate with a wide range of preprint servers and journals that operate on different technological platforms.
d. The conversion process could occupy several different positions in the pipeline of manuscripts, and the preferred position will depend on input from the GB and the ingestion sources. In your response, please describe how you would implement the following options (and any others that you foresee will be compatible with many ingestion sources) and indicate which you see as the preferred option.
    i. All material coming into the CS from ingest sources will be ingested as structured XML. In this case, the conversion tool is offered as a software or hosted as a CS service accessible by API, and individual ingest sources are responsible for implementing it or generating XML by alternative means.
    ii. Material coming from ingest sources can be provided to the CS as the author's original manuscript file (such as .docx or .tex). In this case, the CS would convert the manuscript and contact authors (ie by email) to invite them to proof a rendition of the manuscript after conversion to XML. The ingest source could later retrieve the converted manuscript from the CS.
    iii. The conversion tool could function upstream of ingest sources. A submission tool hosted by the CS could convert manuscripts for author proofing and then provide options for authors to send the converted manuscripts (or original files, or PDFs) to other ingest sources.

5. Automated screening tool

As the volume of preprints in the life sciences continues to grow, more efficient screening mechanisms will be needed. This component will serve to provide an automated way to flag manuscripts that require more in-depth human screening. The expectation is that this tool could be used by ingestion servers, publishers, and other 3rd parties through an API or through their own installation. It could also be used to evaluate the manuscripts coming into the CS from different sources for informational purposes. The CS will not directly perform screening functions (unless a direct ingestion tool is requested by the GB); instead, ingestion sources will be approved by the GB.

In terms of specific requirements, the automated screening tool must support the following functionality:

a. Flag manuscripts with certain characteristics similar to manuscripts that have not passed screening (see "screening training set" in "CS database interfaces for machines"). Initially,

include factors that may include single authorship, presence of keywords, lack of scientific writing style, image style, lack of biological subject matter, presence of human faces, etc.
b. Utilize external services (like plagiarism detection) but ideally develop a viable open source alternative in the long term
c. Provide - via the CS - implementation accessible to machines (a service accessible by API)

6. CS ingest service

The purpose of the ingest service is to import new manuscripts and associated metadata into the CS database. Manuscripts coming into this service will already have been screened by an ingestion source. Conversion to XML could occur upstream or downstream, depending on the implementation of the conversion tool (above).

In terms of specific requirements, the CS ingest service must, as a minimum, support the following functionality:

a. Support the ingest of manuscripts from approved sources (to be determined by the Governance Body) through an API. Recognising that, potentially, some approved ingest sources may not have an API to access their content, suppliers should indicate the approach they would adopt to access and ingest this content.
b. Supports the ingest of the authors' original manuscript file and the converted manuscript in XML (if available - see above) and its appropriate metadata and associated files (figures and supplementary files) or links to these files if too large or more appropriate for a domain-specific repository.
c. Supports the ingest of required metadata including author information (funding source, ORCID, etc as specified by GB), screening history, and manuscript transfer request if applicable.
d. Assigns Crossref preprint DOIs to manuscript and associated supplementary files, if not already provided by the ingestion source.

7. CS direct submission service

Whether the CS should implement a direct submission system remains undetermined at this time and will be further considered at a later date by the Governance Body with more information in hand. For now, please describe how you would implement such a system if it were to be requested and list the costs for this development and operation separately in Section 9 of the application. This service would provide authors with a submission interface - thus providing an alternative to ingesting content from a third party preprint server - and a means to proof their manuscripts after conversion to XML. The direct submission system would also need to make use of both the automated screening service (see Automated screening tool) and human screening. This submission service would then send manuscripts to the ingest service (above).

In terms of specific requirements, the CS direct submission service must, as a minimum, support the following functionality.

a. Provide a direct CS submission interface for human users. This should convert author manuscripts to XML (with proofing) and collect all applicable metadata (including relevant PIDs, journal transfer designation, etc).

b. Provide a human screening service for evaluating content entering through this pathway. This must comply with Governance Body standards. Proceed on the assumption that these will be similar to practices at arXiv, BioRxiv, or PeerJ Preprints.
c. Offer a CS journal transfer system for manuscripts entering the direct submission service
    i. Initially, support FTP transfer
    ii. Later interface with top five manuscript systems (Editorial Manager, eJPress, ScholarOne, Highwire's BenchPress, Aptara PowerReview) in use in life sciences journals with APIs.
    iii. Upon request from author, transfer manuscript to journals.

8. Communication and marketing

a. The above services should be promoted to the community of users (life scientists, developers, funders and institutions, etc) in order to increase awareness and adoption.

2C. Staging

We do not expect that all components listed above will be completed at the same time. While there is flexibility in staging the development of the Central Service, we suggest that features be developed in stages as follows:

- Stage I - milestones to be completed within 6 months of project start

    - Launch of the CS, populated by legacy preprints from approved sources in their original formats (.pdf, .doc, .html, .xml etc.). These files are indexed for full-text search.
    - New manuscripts enter the database by the CS ingest service. Ingestion sources (ie, the preprint servers to which authors submit) must collect original manuscript files (.doc or .tex) which the CS retains for later conversion. However, until the conversion tool is ready, the manuscripts are converted to PDF for display. **No XML is required at this stage.** However, some sources (ie web-based authoring tools) may be able to provide XML, which will be accepted as well.
    - The ingest service also collects required metadata (ORCID, funder IDs, ethical statements, license, a history of how the manuscript has been screened by the ingestion source). These data must be collected from authors by the ingestion source.
    - The ingest service also collects all associated files (supplementary figures, data, movies, etc) and assigns DOIs for any objects that do not already have them, or refers authors to appropriate external repositories if file size limits are exceeded.
    - Ingestion sources are requested to provide all manuscripts regardless of whether they passed screening. While rejected manuscripts will not be displayed to users, they will be later used to train screening systems (possibly anonymized).
    - The CS offers a web interface for browsing and searching manuscripts. Search results display text that has been extracted from PDFs to show search terms in context. If manuscripts are displayed, they can be displayed as PDFs. All pages are tagged with schema.org meta tags to ensure that content is discoverable, and usage metrics are openly shared. The full text may be displayed if requested by the GB.
    - All manuscripts are accessible by bulk download.
    - Researchers are able to sign up for alert services to be notified by email (or RSS or Atom feed or other notification system) of new manuscripts relating to search terms of interest.

- Stage II - milestones to be completed 18 months after project start

    - Manuscript conversion tool is operating. All newly received preprints are converted to XML around the time of submission (by one of the options described above in Manuscript conversion tool). Older preprints for which a compatible authors' file is available will be converted as well.
    - APIs for machine access to manuscripts are available, enabling 3rd party services to request metadata or files from the database.

- Stage III - milestones to be completed 3 years after project start

    - An automated manuscript screening system is in place. Trained by the corpus of both accepted and rejected manuscripts submitted to the CS, this tool will flag manuscripts that are likely to require careful human screening. Like the manuscript conversion tool, this should be easy for ingestion sources to use, either locally or by API calls.

2D. Expected manuscript volumes

Expected manuscript volumes per year are provided below. While submissions cannot be known with certainty and actual numbers might be lower, we include them here because the bidder should have the capability of handling the described volume of submissions.

Year 1 (2017-2018) - 25,000 manuscripts

Year 2 (2018-2019) - 50,000 manuscripts

Year 3 (2019-2020) - 100,000 manuscripts

Year 4 (2020-2021) - 200,000 manuscripts

Year 5 (2021-2022) - 400,000 manuscripts

2E. Support

All software developed for this project must be open source (available on a code-sharing site like GitHub as well as in regular stable releases archived in suitable long-term repositories) and licensed under an OSI-approved license. To encourage reuse, it must be properly documented and supported. This is especially important for the manuscript screening and conversion components. The end-user facing services (discovery, submission, etc.) should also have a helpdesk capable of resolving issues as described below in Contract.

2F. Branding

The bidders of this proposal are able to suggest a new name for the Central Service. The branding of the Central Service would be provisional and requires approval of the Governance Body.

3. Application and Review procedure

3A. Timeline and stages of procurement

February 13	Issue of RFA (this document)
February 24	Q&A session by videoconference, 11am EST - link to join (will be recorded & published)
March 29	Q&A session by videoconference, 9pm11am EDT - link to join (will be recorded & published)
April 23	Last day to submit questions to jessica.polka@asapbio.org
April 30	**Complete proposals must be received via email to jessica.polka@asapbio.org by 5pm EDT**
May	The review of proposals will occur in May. The evaluation committee may ask for a representative to visit and present the proposal in person. ASAPbio will pay for travel if that is requested.
June 30	ASAPbio will identify preferred supplier(s) and present their proposal, along with a proposal for the Governance Body, to the funders' consortium by June 30
September	ASAPbio does not have complete control of the timing of the next steps, although we hope that we will receive a notice of funding and amount of funding in September.
September or October	After a firm funding commitment, ASAPbio will begin discussions of developing a formal contract with the service provider(s) of the successful RFA in September or October.

3B. Submission instructions

All materials must be sent to Dr. Jessica Polka, Executive Director of ASAPbio, by email to jessica.polka@asapbio.org by 5 pm on April 30, 2017. The subject line of the email should read "ASAPbio Central Service RFA response". If you are applying as a consortium of organizations, please send the materials for the entire application together from the point of contact of the lead organization (see The Application).

All materials should be provided in English. Acceptable formats include .doc, .docx, .odt, .tex, .xml, .html, .xls (for budget) and .pdf.

Bidders are encouraged to make all or part of their proposal public, but this will not affect consideration of the proposal.

3C. Evaluation of proposals

Bidders may be excluded from consideration based on legal and financial standing of their organizations. ASAPbio staff, board members, and advisory board members will not be involved in the selection process. However, these groups will identify independent reviewers. These reviewers will be required to complete publicly available COI statements. The names of the review committee members will be released after the submission deadline has passed. Reviewer comments will be made available to the bidders and can be made public with the consent of the bidders in question, but this is not required. Proposals and reviewer comments will be shared with the funder consortium. Proposals will be evaluated on the basis of several criteria.

1. Commitment to the funder principles
   Based on provided answers, the extent to which the bidders' proposal will be compatible with community governance, open source software, transparency, and the view that preprints are global public goods.

2. Technical capability
   Based on organizational history and provided answers, the extent to which the proposal demonstrates an understanding and awareness of the technical requirements in developing and operating the service. The robustness and scalability of the proposed infrastructure will also be a key consideration.

3. Anticipated manuscript volume
   A function of existing or projected manuscript volume, from existing preprint servers or new partnerships with publishers.

4. Administrative capability
   Based on organizational history and provided answers, evidence of the bidders' understanding of the management issues involved and ability to work in partnership to manage and deliver a successful project. Staff competencies and ability to meet defined performance targets will also be taken into account.

5. Previous track record of accomplishment
   Bidders' past record in managing data, software development, manuscript handling, and other factors relevant to this proposal will be considered.

6. Budget
   Value provided in terms of the total cost of implementing, running and developing the service.

Final scores for proposals will be assigned using a process similar to NIH peer review: after discussion, all reviewers will submit an individual impact score for the application. The mean of these scores will form an overall impact score for the application and can be made public with the consent of the respective bidders.

If the evaluators find that none of the applications are suitable, ASAPbio reserves the right to request that bidders who did not collaborate on an application work together to deliver the Central Service. Alternatively, ASAPbio reserves the right to not forward any proposal to the funders consortium and could decide to reissue a RFA.

3D. Policy on information regarding applications

We will be available throughout the process to handle questions from potential bidders, up to the date of April 23, 2017. In the interests of providing a level playing field to other potential bidders, all questions must be routed to Jessica Polka by email (jessica.polka@asapbio.org), irrespective of whether they are procedural, technical or contractual. If questions about the requirements or process of general relevance are received, answers provided by ASAPbio representatives will be posted publicly at asapbio.org/rfa-questions. Questions will be anonymized/summarized before release. To assist with information exchange, we will also hold two bidders' information meetings by video conference on February 24th and March 29th, at 11am EST and 9pm EDT respectively. Recordings of these sessions will be made publicly available at asapbio.org/rfa-questions.

If you are looking for collaborations with other organizations for the purpose of completing a proposal, please leave a message on the public ASAPbio Central Service bulletin board google doc.

3E. Contract

Assuming the Funders' Consortium agree to support the CS, successful bidder(s) will enter into a five year contract with ASAPbio, with break clauses at the end of Stages I & II. Renewal of the contract will depend upon the successful evaluation and recommendation for renewal by the GB and receipt of funds from the funders' consortium. Performance targets include 1) building software, databases, and interfaces described in the yearly contract, and 2) providing services outlined according to the following standards:

- service availability (99.99% uptime, excluding periods of agreed, planned downtime – e.g. for maintenance)
- staffed English-speaking helpdesk (response to all queries within one day, and resolution of 90% of tickets in three days, with holidays and weekends excluded)
- process for escalating unresolved support calls
- server performance (200ms response time, 1s response time for search queries from North America, Europe, South America, and Asia)
- network security (firewalls, DDoS mitigation)
- backup and disaster recovery (ideally, load balanced, redundant data centers enabling failover; entire system is archived at least weekly).

The contract will include a description of service credits that the bidder(s) will offer to ASAPbio in the event that performance targets are not achieved.

4. The Application

1. Information on your organization (one copy per organization)

If this application involves multiple organizations, please fill out separate forms for each organization. Please indicate a lead organization for the purpose of primary ASAPbio contact.

If applying as a consortium, **is this the lead organization?**	Yes ⬜ No ⬜
Name of organization	
Contact person	
Address	
Email address	
Telephone number	
Website	
Federal Tax ID number	
Tax status (eg 501(c)(3) etc)	
Date of registration	
Registered address (if different from above)	

Is your company a subsidiary of another company? If so, please provide the name, Tax ID and status, and registered office address of the holding or parent company and the ultimate parent.

Has your organization been involved in any court action over the last 3 years? If so, please describe.

Provide a list of **Directors:**

Insurance. Please provide details of your organization's insurance protection regarding professional indemnity cover and any other relevant insurance cover.

Policy number	
Insurer	
Expiration date	
Indemnity value	

Financial turnover. Please indicate your total revenue for your last three fiscal years.

Year 1	
Year 2	
Year 3	

Please provide a brief description of your organisation's **Workplace Health & Safety policy**

Please provide a brief description of your organization's **Equal Opportunities policy**

Declaration

I declare that to the best of my knowledge the answers submitted in this proposal (and in any supporting documentation) are correct and that I have all necessary rights or permissions to share the information contained in this proposal with ASAPbio. I understand that the information will be used in the evaluation process to assess my organization's suitability to provide the ASAPbio Central Service. I also understand that any information contained in this proposal may be used by ASAPbio in the discussion and design of the final model, irrespective of whether this proposal is selected.

Name	
Position	
Date	

The following sections are to be completed once per application (ie, once for an entire consortium). Please respond to all of the queries listed below. You may use this document as a template or reference the questions by their numbers in a new document.

2. Summary (up to 1000 words)

2.1. Provide a summary of your strategy for implementing and managing the CS.

2.2. If you represent a consortium of organizations, please indicate why these partnerships were formed and how they will best serve the execution of the project and the interests of the scientific community.

3. Technical capability (2000 words)

Explain why your organization or consortia of organizations are well-qualified to undertake the development and management of the CS. Please address the questions below:

3.1. What previous data archive solutions have you hosted, managed or developed? Please describe size and complexity. Has your organization undertaken projects of the scope of developing the CS?

3.2. Has your organization(s) undertaken software development projects? If so, please describe the project, how it was executed, and the current state of this software.

3.3. Has your organization(s) undertaken open source software development projects? If so, please describe the project, how it was executed, and the current state of this software as well as any experience with managing community contributions.

3.4. Describe any experience/capacity you have in managing the conversion of files to structured XML and highlight any experience you have in validating documents converted to an XML standard.

3.5. Describe what experience/capacity you have in taking data from a variety of sources to create a single database.

3.6. Describe what experience you have in running user-login and submission systems.

3.7. Describe your experience with building and working with APIs.

3.8. Describe what experience/capacity you have in providing a staffed-helpdesk facility.

4. Organization and management considerations (up to 750 words)

4.1. Please insert a diagram of your organization/consortium's planned structure relevant to this project (e.g. Organization Chart)

4.2. Briefly describe the staff that would be responsible for developing and maintaining the CS. State the numbers and qualifications of staff who could work directly on such a project. Name existing

individuals and their credentials if known. Indicate key personnel that would need to be hired.

4.3. Describe what project management methodology you would use, and what experience you have with this methodology. Also, describe other projects you have managed on behalf of a consortium and how you kept the consortium up-to-date with progress etc.

5. Interactions with the Governance Body and Reporting (up to 500 words)

The successful bidder and manager of the CS will have to engage with a Governance Body (GB), managed by ASAPbio. The CS Provider(s) will file a yearly public report to the GB and Funders. The GB will be making recommendations on the performance of the CS Provider(s) to the funders for the purpose of releasing yearly funds to the CS Provider(s). The GB will also set standards for intake of content to the CS and will set yearly priorities and make recommendations. Models similar to this arrangement include Europe PMC and the Protein Data Bank. Please provide any comments or ideas on the relationship and interactions that you (if you obtain the contract) would like to establish with the GB.

6. Execution and Development (up to 3000 words)

6.1. See Requirements for the Central Service. Please address your strategy to implementing each requirement, referring to the numbering described in that section. If a consortium, indicate which organizations would be responsible for which components.

6.2. Provide a timeline for implementation, elaborating or modifying the one proposed in Staging.

6.3. Describe how you would manage the submission and hosting (or referrals to other databases) of large datasets (including a potential ceiling).

7. Sustainability plan (up to 500 words)

ASAPbio will request funding for 5 years. Full or partial funding may be available on the longer term, but bidders should describe sustainable business models that would be compatible with their organization(s) and with the principles of the Central Service. These might include submission fees or fee-for-service arrangements. In our view, unfavorable business models include those where the provider of the service may have a conflict of interest in providing support for preprints as a public good. The Sustainability Plan is provisional and does not commit the bidder(s), ASAPbio, or the Governance Body to execute the plan provided in this RFA.

8. Communications, marketing, and maintenance (up to 1000 words)

8.1. Describe any relevant past experience in marketing. With reference to other projects you have

managed, describe how you promoted the service to the main stakeholders. Also, highlight what marketing strategies worked well, and what approaches were less successful.

8.2. Describe your marketing strategy for the Central Service.

8.3. Describe any relevant past experience in enforcing content standards in a database. Describe how records violating agreed-upon standards were removed and after what decision processes.

8.4. Describe your strategy for removing or flagging content in the Central Service that violates guidelines set forth by the Governance Body. These are yet to be determined, but will likely include dual use research, plagiarised work, or papers with ethically questionable content.

9. Budget (spreadsheet and up to 1000 word justification)

Provide a detailed annual project budget for 5 years of operation in a spreadsheet assuming the schedule of growth listed in Expected manuscript volumes. Provide a written justification of this budget in less than 1000 words.

Please list costs associated with Component 8, the CS direct submission service, in a separate section, as its implementation (and timing thereof) will be decided by the Governance Body in consideration of the productive development of the preprint ecosystem.

Estimate the cost per gigabyte for hosting figures, videos, large datasets, etc.

11. References (up to 5 names)

Provide the names, titles, and email addresses of up to 5 unassociated individuals (who have not been employed by your organization(s) in the past 3 years) who could be contacted with regard to the ability of your organization(s) to execute the work described in this proposal.

Appendix 1. Key Principles for the Central Service developed by the Funders' Consortium

To maximize the chance of funding, this RFA seeks to develop proposals that adhere to points described in the Funders' "Principles for establishing a Central Service for Preprints". We have reproduced the principles in full here:

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 is 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.

Appendix 2. Introduction and Background Leading to this RFA

Despite many advances in internet technologies, the speed with which new research results are being transmitted from an individual life science laboratory to the world-wide scientific community has decreased over the past couple of decades (https://doi.org/10.1073/pnas.1511912112). This trend has largely been an unintended consequence of how scientists interface with current journal systems. In addition to journals requiring an increasing amount of data for publication, it has become increasingly difficult to publish in a timely manner, and serial rejections by journals are now commonplace. Furthermore, if the paper gets sent for peer review, referees and journals require more and more experiments, and two or sometime three rounds of review have become now commonplace. The overall time between completion of a study and its public dissemination in a journal can range from many months to even years. While peer review improves scientific work, the long time between first submission to publication comes with the stiff penalty of delaying the catalytic effects that ensue from having new knowledge in the public domain. Data, methods and ideas made available to other scientists enables them to incorporate this information into their own work and thus promote innovation. In addition, delays in publication hurt young scientists who need to show evidence of productivity in order to obtain PhD degrees and make transitions to subsequent jobs.

While journals provide very valuable services of peer review (a critical part of academic culture and rigor), it is important to develop other mechanisms that can work in parallel and enable faster and more direct scientist-to-scientist communication of research findings. Preprints provide such a mechanism. A preprint is a complete scientific manuscript that is uploaded by the authors to a public server. After a quality control check to ensure that the work is scientific in nature, the posted scientific paper can be viewed without charge through the internet. Thus, a preprint server can facilitate the direct and open delivery of new knowledge to the world-wide scientific community, before traditional validation through peer review.

The preprint server arXiv.org, developed in 1991, has been widely used in physics, mathematics and computer sciences for over two decades, and provides a well-validated model for preprint communication. Physicists, who also generally submit their work in parallel to a journal for peer review, wake up in the morning and read the new postings in arXiv to learn about what is new in the field. Recently, preprint servers have emerged for the life sciences as well, notably bioRxiv (operated through Cold Spring Harbor Laboratory), PeerJ and F1000 Research (where the preprint posting becomes coupled to peer review).

Preprint use by the biology community has been modest until recently, perhaps largely because of lack of awareness and information on this form of direct scientist-to-scientist communication by most life sciences researchers. However, beyond this lack of awareness, some have speculated that preprints will not work in biology, citing that the culture of biology differs from physics/mathematics or that leading scientists and funding agencies in biology will not embrace publications that do not have associated peer review.

1A. ASAPbio and the February 2016 Meeting

ASAPbio emerged as a grassroots organization of four scientists who wanted to test the notion of whether preprints might play a larger role in the life sciences. The first test was a meeting held at the Howard Hughes

Medical Institute in Chevy Chase, Maryland on Feb. 16 and 17, 2016 which brought together ~75 diverse participants, including junior and senior working scientists, experts on data management, and representatives of public and private funding agencies, industry, and scientific journals. In addition to talks, the meeting involved a series of breakout sessions to explore some of the most crucial issues facing preprints, including whether they constitute priority of discovery, should be used as evidence of productivity in grant and promotion reviews, and can be posted without jeopardizing subsequent journal submission. After the discussion, many of these issues were put to an anonymous onsite web vote. Overwhelmingly, the diverse audience voted in favor of preprints and also voiced their opinions in open discussion that preprints could play a very valuable role in scientific communication in the life sciences, as in physics. A summary of this meeting was written by 25 attendees and published in May 2016 (https://doi.org/10.1126/science.aaf9133). The full documentation of the meeting is also available on the ASAPbio web site.

1B. Funders' Workshop, May 2016

One of the most enthusiastic groups of attendees at the February HHMI meeting were the funding agencies. Based upon the success of this meeting, four funding agencies – the Simons Foundation, the Laura and John Arnold Foundation, the Gordon and Betty Moore Foundation, and the Alfred P. Sloan Foundation – funded ASAPbio to continue its efforts in promoting and developing preprints. This funding enabled ASAPbio to hire its first Executive Director, Dr. Jessica Polka. To further probe the interest of funding agencies in preprints, ASAPbio organized a second meeting on May 23-24 at the National Institutes of Health that brought together representatives from 16 funding agencies, a small group of junior and senior scientists, and representatives from preprint servers. The topics included discussion of preprint infrastructure, governance models, citation of preprints in grants, and also the possibility of developing a consortium of international funders to finance the development of new preprint infrastructure and technology. The full meeting agenda is available on the ASAPbio web site. After listening to the discussion, the funders held a two-hour private meeting and continued conversation in the weeks following the close of this meeting. 20 attendees representing many funding agencies then subsequently issued their summary and conclusions. In essence, the consensus was "high enthusiasm about further development of a preprint service for the life sciences." The three specific conclusions and action items were:

1. *" A preprint policy that is as homogeneous as possible across funders is desired, especially in the way that preprints are considered as part of proposal grant submission and review. A subgroup of funders will draft a concept paper addressing some of the policy issues that might arise when implementing such a preprint policy. This draft will be shared with other funders for their input.*
2. *The funders asked ASAPbio to develop a proposal describing the governance, infrastructure and standards desired for a preprint service that represents the views of the broadest number of stakeholders. The proposal should include a budget, goals, milestones and implementation timeline to bring an appropriate community defined preprint service into operation.*
3. *This letter be distributed as widely as possible to inform all stakeholders of the continued interest by funders in expanding the use of preprints by the life sciences community."*

1C. Technical Workshop

Following the guide of these funding agencies regarding point 2, ASAPbio took a next step forward to consider 1) a governance model for preprints, and 2) the infrastructure and standards desired for a preprint service. Regarding governance, a 25 person task force was established in the fall of 2016 to develop a model for governance that reflects the interest of the scientific community and best serves the public good. This process is still ongoing and a report has not been finalized at the time of issuing this RFA. We expect a report and proposal for a governance model to be issued by April, and it will be available for public comment on the ASAPbio web site before it is finalized. Regarding issue 2 (infrastructure and standards), ASAPbio held a Technical Workshop on August 30 at the American Academy of Arts and Sciences. Details of this meeting are again available on the ASAPbio web site and a full report of this meeting was recently published (https://doi.org/10.3897/rio.3.e11825). A major focus of this meeting was the creation of a Central Service for preprints and a discussion of its features and standards.

The report from the technical workshop contains a set of principles and recommendations, which are reproduced here:

General principles	Recommendations
• Preprints are meant to facilitate and accelerate scholarly communication. • Preprint services should encourage open science best practices. • Meet researchers where they are now. Accommodate existing workflows and formats while moving toward best practices over time. • Remember the motivations of researchers (including credit, career progression, and convenience). • Take advantage of available technology. Preprint technology should be built quickly in a way that can be extended and expanded in the long term by many parties. • Allow preprints to be transferred to journals in	• Focus on standards. Use schema.org compatible meta-tags and recognized API standards such as OAI-PMH or equivalent. Use the standard persistent identifiers adopted by the community so that we can systematically link up resources, people, and organizations. For example, include person identifiers, document identifiers, identifiers for data, etc., and authenticate them to the extent possible. • Make markup consistent. Engage with JATS4R or similar initiatives and follow existing recommendations on tagging. • Develop open technologies. Permissive, open licenses on software should be strongly encouraged, and serve as the default for new code written for any ASAPbio projects. • Encourage best practices for screening. Manuscripts must be screened by humans before posting, and takedown policies need to be implemented in a standardized fashion. • Stay simple. Accept submissions in Word format and LaTeX and display them in PDF from day 1. The originally submitted files should also be retained and made accessible for mining and processing. • Support open source conversions. Request and support the creation of an open-source document conversion tool from popular formats like Word and LaTeX to consistent markup (JATS and/or XHTML). • Develop machine screening algorithms. To learn from the process, require all manuscripts (accepted and rejected)

formats that fit journal workflows.	to be collected along with their screening status to form a database of content; use this to improve machine screening algorithms. ● Streamline transfers. Support simple transfer of articles to traditional journal workflows. ● Promote data sharing. The service should make it easy for authors to refer readers to data, software and other relevant materials. Encourage and facilitate deposition of data in appropriate repositories. ● Directly accommodate deposition of supplementary files (such as figures, movies, and text), which should be given their own unique identifiers and be preserved and indexed appropriately.

While opinions varied on some issues, the concept of the Central Service met with broad support and now is the subject of this RFA. The overall objective of the Central Service is to serve as a database that aggregates preprint content from several sources; provides unified standards for intake in terms of quality, ethical compliance, and metadata; stores the data in a reliable manner; and provides APIs that allows third parties to access and use this unified database. The need for a new entity (the Central Service) results from the fact that preprint content in the life sciences, unlike physics, is likely to come from multiple sources. Already, several preprint servers in biology exist. In addition, journals may enter into this space and make their submissions, with author approval, available as a preprint. Thus, preprint content will likely grow from multiple sources, emphasizing the need to collect and provide uniformity and ease of access to this relatively new mechanism for communication in the life sciences.

Furthermore, the utility of preprints in the life sciences is limited by their availability in a format (PDF) that is difficult for humans and machines to use. The Technical Workshop reinforced the idea (originally discussed at the Funders' Workshop) that converting preprints into better-structured formats such as XML would enable them to be readily passed as data to software tools that promote new kinds of scientific discovery. This development would bring preprints into the modern era of information.

1D. Interactions with Funding Agencies Toward the Development of a Central Service

Subsequent to the Technical Workshop, ASAPbio contacted the funding agencies to discuss how to bring the Central Service from concept into reality. Ten funding agencies participated in in-depth discussions of whether they, in principle, might be willing to support a Central Service as a consortium and what type of Central Service they would like to see develop and be willing to support. ASAPbio was not involved in these discussions, and their issued statement on the Central Service reflects their independent deliberations. The request of the funders to ASAPbio is stated below:

"While the Funders... 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."

In keeping with this request, ASAPbio is now issuing this RFA. After receiving responses to this RFA, an independent evaluation panel (described in Evaluation of proposals) working on behalf of ASAPbio will select a service provider(s) to develop and operate the CS. ASAPbio will then construct a single proposal that will include operating expenses for 1) developing and operating the CS, 2) operations and meetings of the Governance Body (GB) for the CS, and 3) operating expenses so that ASAPbio can manage funds and communications from the funding agencies, organize the Governance Body, and provide information to the community about preprints and the operations of the CS and GB. We intend to approach the funding consortium in June with this unified proposal. We cannot guarantee that the funders will decide to fund the proposal that we put forward. Thus, even if a proposal to this RFA is selected, there is no guarantee that it will be funded. We also cannot project the total budget that would be available (see Timeline and stages of procurement).

Acknowledgments

We are grateful for many helpful comments provided by the signatories of the funder principles, ASAPbio board members, and Daniel Mietchen.