ASAPbio Preprint Platform Directory: 2019 data

In collaboration with Jamie Kirkham & Fiona Murphy

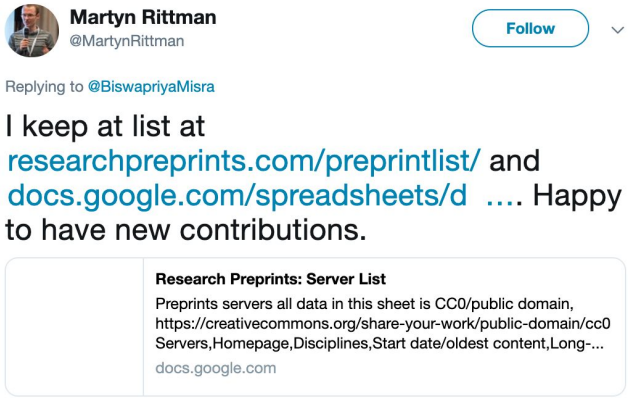
Naomi Penfold, ASAPbio

Jan 20, 2020 #bioPreprints2020

ASAPbio

Many preprint platforms…

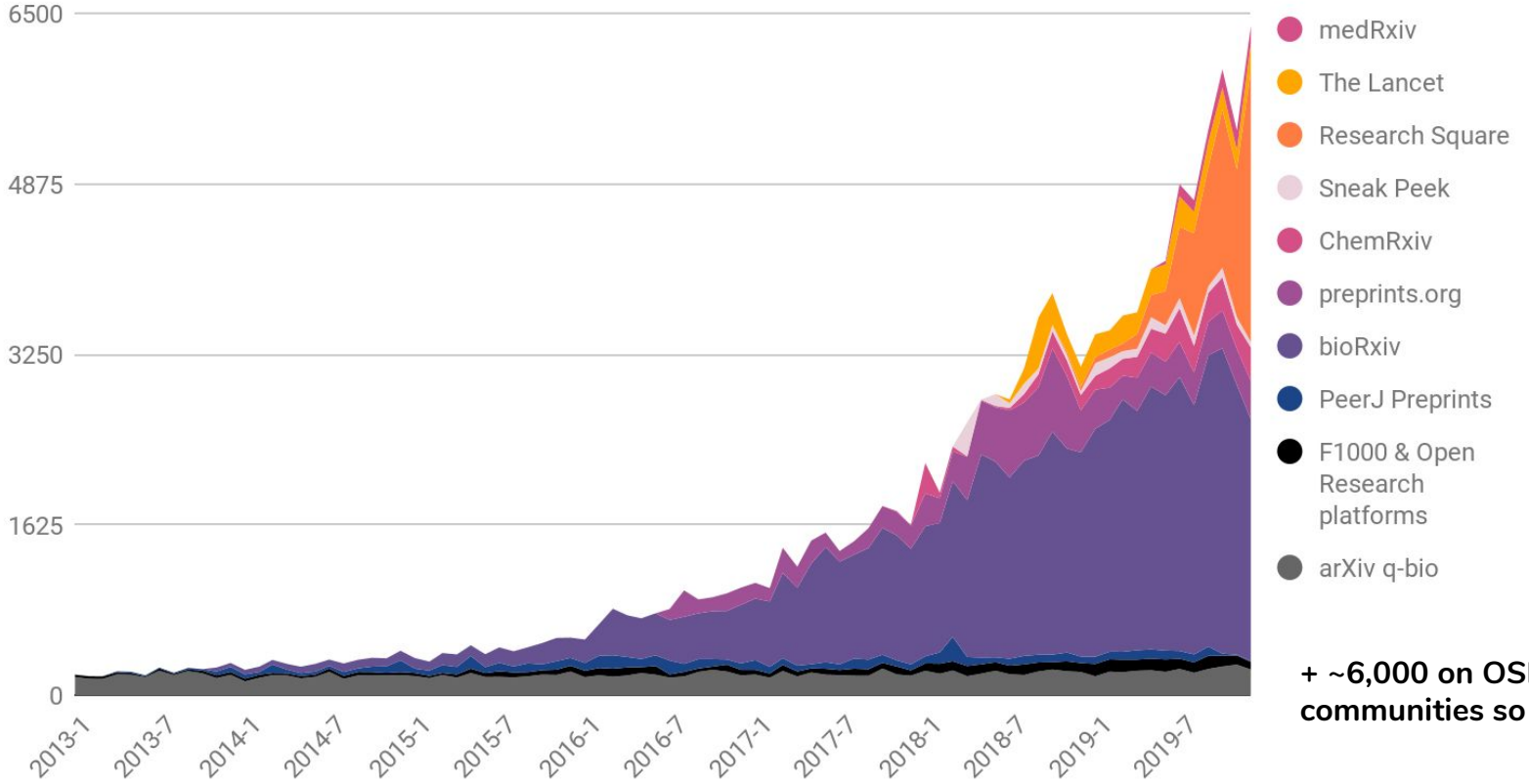
Updated from image compiled by Jeroen Bosman (twitter:@jeroenbosman) via Bianca Kramer (twitter:@MsPhelps)



https://twitter.com/MartynRittman/status/907135259854155776

Biomedical preprints per month

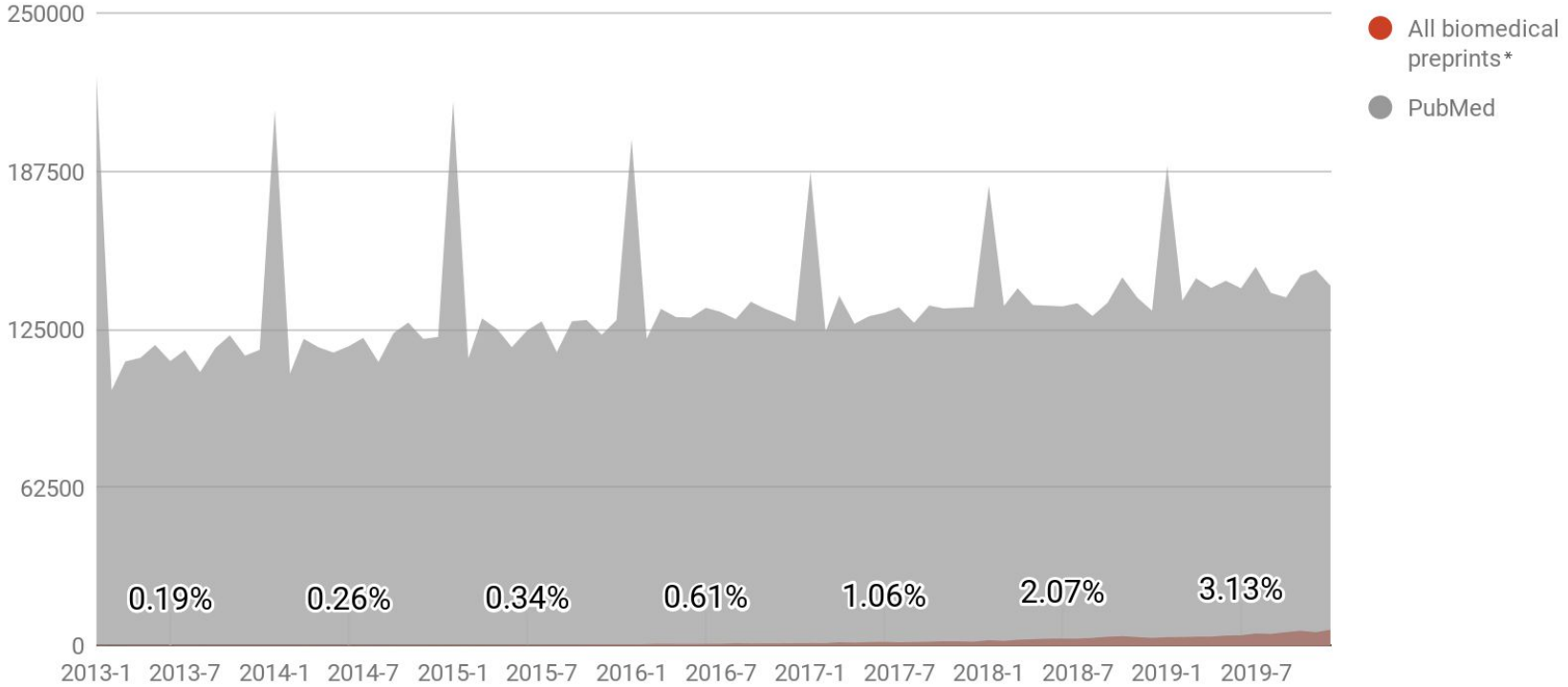
Sources: Jordan Anaya (PrePubMed), Naomi Penfold, EuropePMC, Research Square, SSRN



+ ~6,000 on OSF Preprints & communities so far

Papers per month

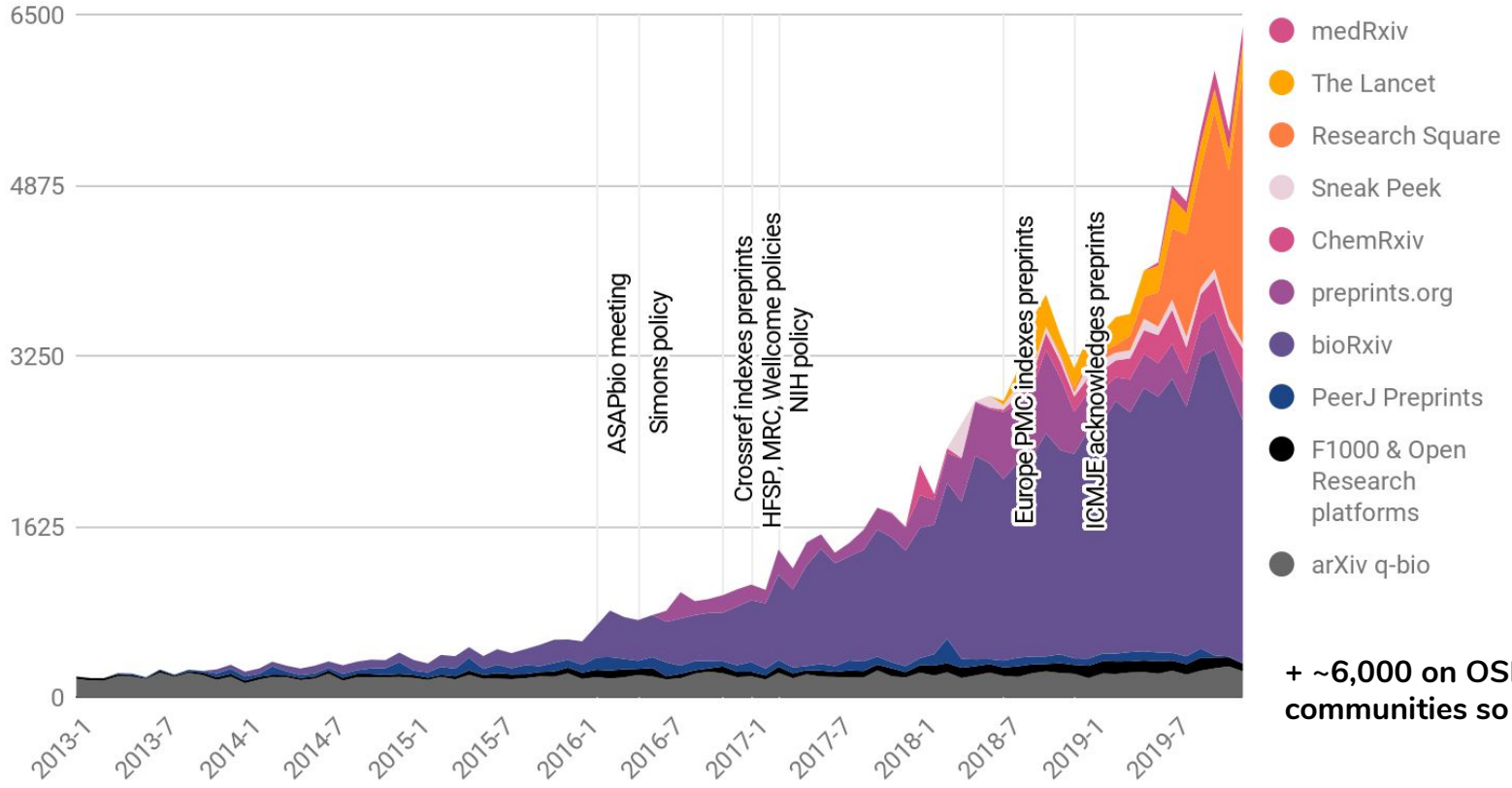
Labels reflect percentage of preprints/PubMed papers as a yearly averages



* Not including ~6,000 on OSF Preprints & communities so far

Biomedical preprints per month

Sources: Jordan Anaya (PrePubMed), Naomi Penfold, EuropePMC, Research Sqaure, SSRN



+ ~6,000 on OSF Preprints & communities so far

Explicit guidance from funders

Guidance for selecting interim research product repositories

Interim research products rely on repositories to make them public. The repository market is growing rapidly, and in many scientific disciplines, norms for interim research products are still evolving.

The NIH would like to ensure that practices for interim products facilitate the impact, measurement and the integrity of the scientific record. Specifically, the NIH strongly encourages interim research products arising from NIH funds to be deposited in repositories that ensure:

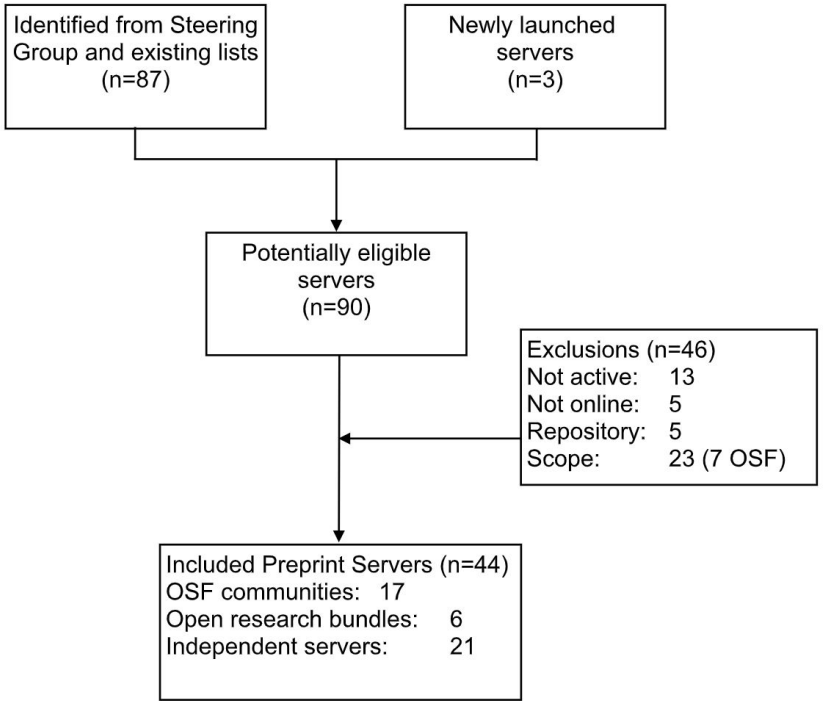
- Content is findable, accessible, interoperable and reusable.
- Interim product metadata, including usage statistics, are open, and easy to access by machines and people (e.g. via application program interfaces).
- Content is easy to use by machines and people. This access is both a function of permission (e.g. use of Creative Commons licenses) and technology (e.g. application program interfaces).
- Policies about plagiarism, competing interests, misconduct and other hallmarks of reputable scholarly publishing are rigorous and transparent.
- Records of changes to the product are maintained, and users have clear ways to cite different versions of the product.
- Links to the published version, if available.
- A robust archiving strategy that ensures long-term preservation and access.

NIH Notice NOT-OD-17-050:
https://grants.nih.gov/grants/guide/notice-files/not-od-17-050.html

Funder principles & technical recommendations published after previous workshops

Where are we now?

44 platforms surveyed



JJ Kirkham et al, *in preparation*

2. **Platform's description of the screening check**
Quote policy and include URL where possible; If none, state "none", or enter "unknown - information not found online" or "unknown - question not considered"

3. **Who screens submissions?**
Enter details found online about who is involved in the screening process, including URL(s) where appropriate; or enter "unknown - information not found online" or "unknown - question not considered"

4. **Are researchers involved in screening process?**
Mark only one oval.
- Yes
- No
- Unknown - information not found online
- Unknown - question not considered

5. **Screening checks - overview**
Are the following checked for during screening. More detailed information in following questions.
Check all that apply.
- Competing interests declared
- Plagiarism (e.g. run through software)
- Misconduct or integrity checks
- Authors are genuine
- All authors provide permission to post
- Compliance with ethical and legal requirements
- Funders acknowledged
- Clinical trial registration
- Other:

28. **Metadata availability**
How is metadata made available to third parties; if unknown, enter "unknown - information not found online" or "unknown - question not considered"
Mark only one oval.
- Openly available direct from the platform via API
- Openly available via Crossref
- Available for free upon request
- Available with charge upon request
- Not available
- Unknown - information not found online
- Unknown - question not considered
- Other:

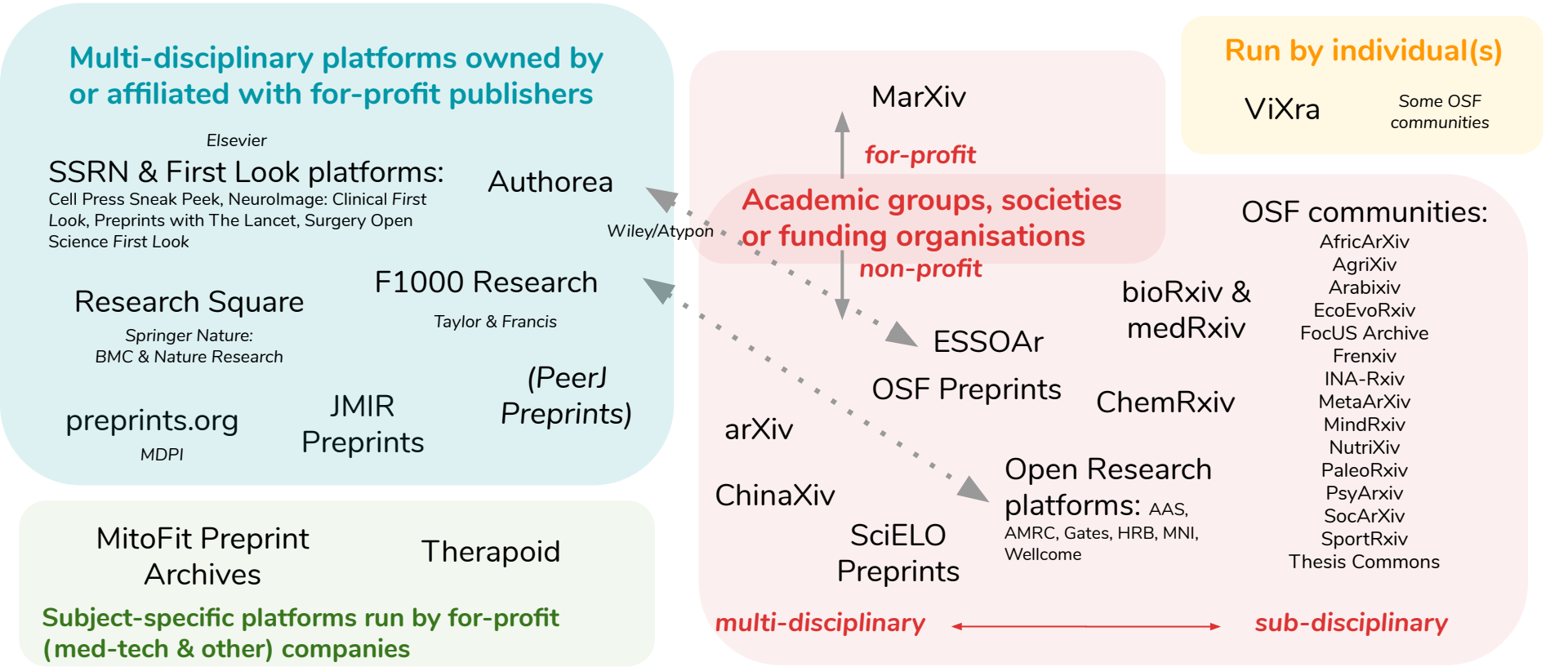
29. **What does metadata contain?**
Select any options where the server indicates they share it in the metadata (even if the data field is not always complete or accurate).
Check all that apply.
- Title
- Identifier (e.g. DOI)
- Publication/deposition date
- Author name(s)
- Abstract
- Relational link to final journal publication (e.g. in crossref metadata)
- Author affiliation(s)
- Funder acknowledgement(s)
- License type(s)
- Subject category
- Full-text content
- References
- Unknown - information not found online
- Unknown - question not considered
- Other:

30. **Is Crossref metadata record updated when preprint is linked to publication (when known)?**
This affects how europePMC and other Crossref metadata users can index preprints and provide third-party services.
Mark only one oval.
- Yes
- No
- Other:

110 survey questions or fields completed through online research where available, verified and updated in consultation with platform representative, where possible; original data simplified and verified

Landscape of platforms

- Access to money, staff, time, publishing know-how
- Philosophy on amount of gatekeeping versus speed & transparency
- Motivations: from publisher-driven preprints to publishing-disruptive preprints



Where are we at?

Previous recommendations

Unitalicized recommendations: Funder principles (2017)

Italicized recommendations: *Technical workshop (2017)*

& NIH guidelines

User friendliness

- *Stay simple. Accept submissions in Word format and LaTeX and display them in PDF from day 1.*

- *Streamline transfers. Support simple transfer of articles to traditional journal workflows.*

B2J, MECA, In Review, SSRN First Look

- be easy to use, with rich search and discovery tools and visible usage metrics

EuropePMC

- aim to attract widespread community support

See monthly growth

- posting and commenting should be incentivised

TRiP

Screening and ethics

• *Manuscripts must be screened by humans before posting, and takedown policies need to be implemented in a standardized fashion.*	**Variable**
• be accompanied by clear funding acknowledgements, declaration(s) of competing interests	**Variable**
• adhere to standard scholarly publication practices, including ethical, legal and societal standards;	**Variable**

What checks does a preprint submission go through before it is posted online?

(+) indicates screening process involves active researchers

Submission screening checks	AfricArxiv, Arabixiv, MindRxiv (+), OSF Preprints, Thesis Commons, Authorea, SSRN	MetArXiv, SportRxiv (+)	PaleorXiv (+)	PsyArXiv	ChinaXiv (+)	MarXiv *(no longer on OSF, still on Zenodo)*	ViXra *(no formal checks, but submissions rejected for reasons relating to:)*	arXiv (+)	bioRxiv (+)	ESSOAr	Frenxiv (+)	MitoFit Preprint Archives (+)	Preprints.org (+)	PeerJ Preprints *(closed)*	EcoEvoRxiv (+)	Preprints with The Lancet	ChemRxiv (+)	SciELO Preprints *(when fully launched)*	Open Research platforms, F1000 Research	Research Square	MedRxiv
Scope (scientific, not spam, relevant material, language)	✓	✓	✓	✓			format (PDF)	✓	✓	✓			✓		✓	✓			✓ (not said for ORs)	✓	✓
Author(s) genuine (& eligible, if needed)		✓			✓			✓		✓	✓	✓	✓				✓	✓	✓		
Plagiarism			✓	✓	✓	✓	✓	✓	✓	✓			✓ (as needed)		✓		✓	✓	✓	✓ (only for BMC submissions)	✓
Legal/ethical/societal issues & compliance				✓		✓	✓	✓	✓		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
Misconduct or integrity checks					✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		✓	✓			✓
Author(s) consent										✓							✓	✓	✓		
COIs declared (or none)											✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
Funder acknowledgement												✓		✓	✓	✓	✓	✓	✓	✓	✓
Clinical trial registration																✓	✓	✓	✓	✓	✓
Data/code availability													✓				✓	✓	✓		
Minimal reporting standards																			✓		✓
Is it dangerous to human health?																				✓	✓
Time taken					?																

Up to 1 day | Up to 2 days | A few days | Up to 1 week | Up to 2 weeks

Have checks, details not known: FocUS Archive, SocArxiv

No checks: agriXiv, Therapoid (rely on moderation)

No information regarding checks: INA-Rxiv, NutriXiv, 'First Look' platforms (Cell Press Sneak Peek, NeuroImage: Clinical First Look, Surgery Open Science: First Look), JMIR Preprints

Derived from Penfold, Naomi C, Murphy, Fiona L M, & Kirkham, Jamie J. (2020). Practices and policies of preprint platforms for life and biomedical sciences (Version 1.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.3612693

Submission screening checks	Some OSF-based platforms, Authorea, SSRN	Some OSF-based platforms	ChinaXiv (+)	arXiv (+)	bioRxiv (+)	ESSOAr	Preprints.org (+)	EcoEvoRxiv (+)	ChemRxiv (+), SciELO Preprints (when fully launched)	Open Research platforms, F1000 Research	Research Square	MedRxiv
Scope (scientific, not spam, relevant material, language)	✓	✓		✓	✓	✓	✓	✓		✓ (not said for ORs)	✓	✓
Author(s) genuine (& eligible, if needed)		✓	✓	✓		✓	✓		✓	✓		
Plagiarism		✓	✓	✓	✓	✓	✓ (as needed)	✓	✓	✓	✓ (only for BMC subs)	✓
Legal/ethical/societal issues & compliance		✓		✓	✓		✓	✓	✓	✓	✓	✓
Misconduct or integrity checks		✓	✓	✓	✓	✓		✓	✓			✓
Author(s) consent						✓			✓	✓		
COIs declared (or none)		✓						✓	✓	✓	✓	✓
Funder acknowledgement								✓	✓	✓	✓	✓
Clinical trial registration									✓	✓	✓	✓
Data/code availability							✓		✓	✓		
Minimal reporting standards										✓		✓
Is it dangerous to human health?											✓	✓
Time taken			?									

(+) indicates screening process involves active researchers

Up to 1 day | Up to 2 days | A few days | Up to 1 week | Up to 2 weeks

Derived from Penfold, Naomi C, Murphy, Fiona L M, & Kirkham, Jamie J. (2020). Practices and policies of preprint platforms for life and biomedical sciences (Version 1.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.3612693

Are policies & practices clear?

Guidance for selecting interim research product repositories

Interim research products rely on repositories to make them public. The repository market is growing rapidly, and in many scientific disciplines, norms for interim research products are still evolving.

The NIH would like to ensure that practices for interim products facilitate the impact, measurement and the integrity of the scientific record. Specifically, the NIH strongly encourages interim research products arising from NIH funds to be deposited in repositories that ensure:

- Content is findable, accessible, interoperable and reusable.
- Interim product metadata, including usage statistics, are open, and easy to access by machines and people (e.g. via application program interfaces).
- Content is easy to use by machines and people. This access is both a function of permission (e.g. use of Creative Commons licenses) and technology (e.g. application program interfaces).
- Policies about plagiarism, competing interests, misconduct and other hallmarks of reputable scholarly publishing are rigorous and transparent.
- Records of changes to the product are maintained, and users have clear ways to cite different versions of the product.
- Links to the published version, if available.
- A robust archiving strategy that ensures long-term preservation and access.

NIH Notice NOT-OD-17-050:
https://grants.nih.gov/grants/guide/notice-files/not-od-17-050.html

F1000 platforms, MitoFit Preprint Archives, PeerJ Preprints (closed), Preprints.org – have all these policies described online

Most have plagiarism policy described online

Some platforms have policies but not online

Some have no policies

If content is withdrawn, it retains web presence for some platforms – this is not standard across all platforms

Are archiving practices clear?

Guidance for selecting interim research product repositories

Interim research products rely on repositories to make them public. The repository market is growing rapidly, and in many scientific disciplines, norms for interim research products are still evolving.

The NIH would like to ensure that practices for interim products facilitate the impact, measurement and the integrity of the scientific record. Specifically, the NIH strongly encourages interim research products arising from NIH funds to be deposited in repositories that ensure:

- Content is findable, accessible, interoperable and reusable.
- Interim product metadata, including usage statistics, are open, and easy to access by machines and people (e.g. via application program interfaces).
- Content is easy to use by machines and people. This access is both a function of permission (e.g. use of Creative Commons licenses) and technology (e.g. application program interfaces).
- Policies about plagiarism, competing interests, misconduct and other hallmarks of reputable scholarly publishing are rigorous and transparent.
- Records of changes to the product are maintained, and users have clear ways to cite different versions of the product.
- Links to the published version, if available.
- A robust archiving strategy that ensures long-term preservation and access.

NIH Notice NOT-OD-17-050:
https://grants.nih.gov/grants/guide/notice-files/not-od-17-050.html

Almost all have a preservation plan for read-access (or plan to implement): Portico, COS's fund, mirror sites, dark archives or local backups

This is not transparent to researchers or readers

Interoperability & reuse

- be usable for text and data mining (TDM), ideally with full-text available via APIs, and with permissive licensing, which is explicitly stated in a machine-readable way;
- *the preprinting ecosystem should be interoperable and leverage existing standards.*
- *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."*

Full-text available from F1000 Res by request, other platforms do not provide; licensing: some require CC-BY, many provide author choice

ORCIDs

XML: bioRxiv, F1000 platforms

Metadata contents & accessibility	arXiv	Authorea	bioRxiv	ChemRxiv – not verified	ChinaXiv	ESSOAr	F1000 Research & Wellcome Open Research	Other Open Research Platforms	JMIR Preprints – not verified	MedRxiv	MitoFit Preprint Archives (+) – not verified	OSF Preprints & communities - not verified	PeerJ Preprints (closed)	Preprints.org (+) - not verified	Research Square	SciELO Preprints (when fully launched)	SSRN	Therapoid	ViXra - not verified
Title										✓									
Abstract										✓									
Publication or deposition date										✓									
Author name(s)										✓									
DOI or identifier	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
Subject category		✓	✓	✓	✓	✓	✓			✓	✓	✓	✓		✓	✓		✓	
Author affiliations(s)		✓	✓	✓		✓	✓			✓	✓	✓	✓		✓	✓			
License			✓	✓		✓			✓	✓	✓	✓	✓	✓	✓	✓			
Relational link to journal	✓		✓			✓				✓		✓	✓	✓	✓				
Funder acknowledgement(s)				✓		✓	✓			✓	✓		✓		✓				
Full-text content		✓	✓	✓			✓								✓				
References			✓	✓						✓					✓		✓		
Other		Keywords				ORCID, keywords	Competing interests, Referee (name, affiliation, report), Commenter (name, affiliation, comment text), Study type, Article type; and for F1000 Research: Article is a Faculty Review (yes/no)		Journal currently submitted to		ORCID if provided, Language, Version					Language		Anything else the author provides'	
Openly available via:	API	?	API	API	API	Crossref	API	Upon request	?	Crossref	DataCite	API (mostly)	Crossref	Crossref	Crossref	API	?	Crossref	?

Derived from Penfold, Naomi C, Murphy, Fiona L M, & Kirkham, Jamie J. (2020). Practices and policies of preprint platforms for life and biomedical sciences (Version 1.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.3612693

Preprint-level usage metrics in metadata at source or elsewhere?

Guidance for selecting interim research product repositories

Interim research products rely on repositories to make them public. The repository market is growing rapidly, and in many scientific disciplines, norms for interim research products are still evolving.

The NIH would like to ensure that practices for interim products facilitate the impact, measurement and the integrity of the scientific record. Specifically, the NIH strongly encourages interim research products arising from NIH funds to be deposited in repositories that ensure:

- Content is findable, accessible, interoperable and reusable.
- Interim product metadata, including usage statistics, are open, and easy to access by machines and people (e.g. via application program interfaces).
- Content is easy to use by machines and people. This access is both a function of permission (e.g. use of Creative Commons licenses) and technology (e.g. application program interfaces).
- Policies about plagiarism, competing interests, misconduct and other hallmarks of reputable scholarly publishing are rigorous and transparent.
- Records of changes to the product are maintained, and users have clear ways to cite different versions of the product.
- Links to the published version, if available.
- A robust archiving strategy that ensures long-term preservation and access.

NIH Notice NOT-OD-17-050:
https://grants.nih.gov/grants/guide/notice-files/not-od-17-050.html

Most display views & downloads to the reader

bioRxiv monthly usage statistics available via API (not at individual preprint level)

Europe PMC displays citation count & altmetric

Data and supplementary files

- [be accompanied by] availability of supporting data (such as data availability statements, permissively licensed data);
- *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.*

DAS: variable; rarely separate licensing for supporting materials

Variable

Support submission of most files (or deposition elsewhere), rarely further support

Tool development

- 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).

No

Next steps

Now: Data available from Zenodo (v1.0, CC0) – DOI: 10.5281/zenodo.3612693

This week: let's identify some practical next steps to build trust and improve discoverability and interoperability!

Coming soon: online searchable directory from ASAPbio

Next: manuscript in preparation (Kirkham et al,...)

Thank you

Prof Jamie Kirkham
Centre for Biostatistics, Manchester Academic Health Science Centre, University of Manchester, UK

Fiona Murphy, DPhil
Associate Fellow, University of Reading, UK

Naomi Penfold, PhD
ASAPbio, San Francisco, USA

With contributions from:

Jessica Polka, ASAPbio funders & advisors

Jamie's Steering Group: Isabelle Boutron, John PA Ioannidis, John Inglis

Martyn Rittman, Michael Parkin

With thanks to all the preprint platform representatives who confirmed and updated their data

ASAPbio Member Advisory Group funds work to advance the productive use of preprinting in the life sciences