Using open science to speed dissemination, reduce burden and measure impact

Neil Thakur, PhD
Office of Extramural Research
NIH

NFAIS
3-1-18
Increase opportunities
Reduce burden
Better incentives and measures?
Preprints?

• Growing recognition that interim research products could speed the dissemination of science and enhance rigor
  ▪ Many disciplines have been using preprints for years (economics, physics)
  ▪ Groups like ASAPbio suggest expanding preprints could increase the impact of NIH research
  ▪ Groups like the Center for Open Science suggest preregistration could enhance rigor of NIH supported research

• Change happening at different rates for different disciplines

• No clear definitions

• NIH rules are narrow; except for the reference section of applications
Interim Research Products

**Broad definition (to avoid many small rules): Complete, public research products that are not final**

**Preprints** are complete and public drafts of scientific documents. Arguably speeds dissemination, establishes priority, generates feedback, and may reduce publication bias.

**Preregistering a protocol** is publicly declaring key elements of a research project in advance. May reduce biases like p-hacking.
NIH’s interest: Stable infrastructure to advance science

Integrity
• Open (accessible, impactful, egalitarian)
• Preserved and reliable (versioning, linkages, metadata)

Encourage responsible standards and beneficial innovation
• Allow disciplines to adopt at their own pace
• Innovation can increase rigor and dissemination
• Prevent bad practices from taking root

Ensure more preprints= better science

NIH policy change: Interim research products can be cited anywhere other research products are cited. (e.g. biosketches, progress reports) NOT-OD-17-050
Integrity through standards and best practices

Citation format standards: DOI, list object type (e.g. Preprint, protocol)


<table>
<thead>
<tr>
<th>To cite preprints arising from NIH awards</th>
<th>Repository best practices</th>
</tr>
</thead>
<tbody>
<tr>
<td>• DOI</td>
<td>• FAIR principles (Findable, Accessible, Interoperable, Reusable)</td>
</tr>
<tr>
<td>• Recommend CC-BY license</td>
<td>• Open metadata</td>
</tr>
<tr>
<td>• Statement: not peer-reviewed</td>
<td>• Machine accessible and readable</td>
</tr>
<tr>
<td>• Acknowledge funding</td>
<td>• Transparent policies</td>
</tr>
<tr>
<td>• Declare competing interests</td>
<td>• Versioning</td>
</tr>
<tr>
<td></td>
<td>o links to published version</td>
</tr>
<tr>
<td></td>
<td>o tracks key changes</td>
</tr>
<tr>
<td></td>
<td>• Permanent/archival plan</td>
</tr>
</tbody>
</table>

Role in NIH review
Possible need and potential impact of citing interim products on peer review of NIH applications.

Impact on Review

Support in Review?

<table>
<thead>
<tr>
<th>Category</th>
<th>N/A</th>
<th>Supportive</th>
<th>Unsupportive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blank/No Response</td>
<td>40%</td>
<td>50%</td>
<td>10%</td>
<td>30</td>
</tr>
<tr>
<td>Other</td>
<td>36%</td>
<td>64%</td>
<td>0%</td>
<td>14</td>
</tr>
<tr>
<td>Publisher</td>
<td>67%</td>
<td>33%</td>
<td>0%</td>
<td>3</td>
</tr>
<tr>
<td>Scientist/Society</td>
<td>45%</td>
<td>32%</td>
<td>23%</td>
<td>22</td>
</tr>
<tr>
<td>Scientist/Author</td>
<td>30%</td>
<td>67%</td>
<td>3%</td>
<td>276</td>
</tr>
<tr>
<td>Scientist/Author; Other</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>2</td>
</tr>
<tr>
<td>Scientist/Author; Publisher</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>1</td>
</tr>
<tr>
<td>Scientist/Author; Publisher;Scientific Society</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>1</td>
</tr>
<tr>
<td>Scientist/Author; Scientific Society</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>33%</td>
<td>63%</td>
<td>4%</td>
<td>351</td>
</tr>
</tbody>
</table>

Cite interim products using proper citation formats

- RPPRs: Preprints in Section C1; other products in section C5
- Applications: Progress Report Publication List, Biosketch
Impact for Reviewers (from December 2016 RFI)

How NIH reviewers might evaluate citations of interim research products in applications

### Advice for reviewers

- **Read the product, yes**: 86
- **Blank/No Response**: 78
- **Treat like other publications**: 44
- **Publications, weaker evidence than**: 38
- **Unpublished data, same evidence as**: 23
- **Treat like other products**: 20
- **Publications, same evidence as**: 18
- **Burden, Reviewer**: 15
- **Importance of citation formats**: 7
- **Read the product, no**: 2

### Reviewer Guidance

- Write simple definitions for interim research products
- Note that these products are not peer-reviewed
- NIH is neutral on whether reviewers should read references
Overlapping problems with research and career information

• **Duplicate data and wasted effort:** Researchers have to curate and combine data that is scattered across public and private sources—ORCID, SCOPUS, PubMed, RPPRs, Vivo, Trellis, etc., and must do this in multiple times in multiple systems.

• **Poor tracking and measurement**— Funders can’t track their impacts on researcher careers, especially across different funders.

• **Inefficient research networks**— Researchers and associated groups do not use modern technology for networking and hiring (e.g., finding mentors, collaborators, employees, reviewers, etc.)

• **Bad incentives**— Current measures of research productivity do not adequately incentivize openness, rigor and impact. Current fragmentation in research and career data and reporting makes it difficult to implement new measures.
Goals for a better impact infrastructure

• Track funder impact

• Encourage development of better productivity measures and incentives

• Support collaboration and networking services

• Maintain researcher control and privacy

• Reduce researcher burden
Working at scale: Design or adoption challenge?

Do funders have the leverage to address many larger goals?
  • Funders are small in scale
  • Research funding is only one of many incentives and systems

Funder systems are not the burden, so silos are not the answer
  • FDP experience with profile data: Fragmentation, burden, inefficiency
How many users does your system have?
Create a comprehensive research impact infrastructure with unique identifiers

**Link**
- Products (RRID, DOI, ORCID)
- Funding (DOIs?)
- People (ORCID)
- Institutions (?)

**Enable**
- Burden reduction
- Impact analysis
- Metrics
- Innovation and economic growth
Persistent identifiers

ORCID
– A persistent unique identifier for researchers
– Helps track and validate people/product associations
– Over 4M users, supported by thousands of journals

Digital Object Identifiers (DOIs)
– Developed as a universal, persistent article identifier to overlay multiple publisher data systems
– Infrastructure for metadata, validation, citation tracking
– Now extended to data and other research objects
– 63M articles, 11M books and book chapters, and many other products
DOIs for funding (grants, contracts, etc)

Utilize the publications tracking infrastructure to track grants

- Better tracking of people across their careers and funding agencies
- More accurate identification of research products
- More robust data to identify potential reviewers and assess conflicts of interest
- Validation for grant /product associations

As an overlay, a universal funding number system for all funding agencies

- Provide a ‘common denominator’ funding identifier format to harmonize NIH’s grants system and contract system, and harmonize with other funders
- An inexpensive way for funding agencies to develop unique identifiers for their funding. Requires permanent location for funding information

https://www.crossref.org/community/funders/
ORCID will enhance their data model and 3rd party service integrations to:
• broaden connections to research and career data usually reported on CVs
• link researchers to funding and professional activities with verified and structured data
• serve as an open hub for other systems
• will also explore institutional identifiers

Goals
• **Reduce researcher burden** of applying for funds and maintaining multiple profiles
• **Track impact** of research and professional development through transparently-curated open data
• **Support collaboration and networking services** to build efficient and equitable markets for reviewers, collaborators, mentors, etc.
• **Maintain researcher control** of their own data and how it is used across platforms
• **Encourage development of better productivity measures and incentives**

https://orcid.org/content/orbit-project
Use Case: Better Measures

ORBIT aggregates by person...
- Products (DOIs, Etc)
- Funding (DOIs?)
- Institutions (institutional identifiers?)

Product level metrics can be retained in metadata
- Relative citation ratio
- Openness? (licenses?)
- Rigor? (badges?)

Product level metrics can aggregate to...
- Person level measures
- Award level measures
- Funding initiative level measures
- Institution level measures
Additional use cases and information
Use Case: Application Forms

- ORCID/ORBIT data hub
- SciENcv writes creates biosketches for NIH, NSF, ED
- User approval for data linkage
- Reduced burden, validation, structured data
- Scaling: eRA as 1/10th users of ORCID
Use case: Better university data

• ORCID/ORBIT integrates data streams for linked accounts
• Users can manage their data in the system they prefer
• Primary source of burden for PI profiles
ORCID integration with NIH systems

ORCID provides investigators with persistent digital identifiers and helps them track their research products

Phase 1: integration with SciENcv
• Link to ORCID in SciENcv and download ORCID citations into biosketches

Phase 2 (current): Allow ORCIDs in eRA profiles
• Facilitate data exchange, funding/ORCID linkages

Phase 3 (future): Expand ORCID data model and integration with eRA
• Use ORCID data to automate other forms like Other Support, RPPR?
• Upload NIH data (funding, products, profile data) into ORCID?
• Use ORCID as a hub and interchange for all profile data, reducing burden for federal and private profile systems?
SciENcv = Science Experts Network Curriculum Vitae

Vision - Let investigators harvest their data from multiple systems to support funding applications, reporting and collaboration with less burden and complexity

Goals

- **Reduce burden** of applying for federal funds and maintaining federal profiles
- **Track impact** of federal investments in science and scientist careers through scientist-curated data
- **Support collaboration and networking services** to find reviewers, collaborators, mentors, etc.

Products to date

- NIH biosketches, NSF biosketch, Ed IES biosketch
- Embedded XML
- Integration with ORCID, Fastlane, PubMed and eRA
- Bulk upload of citations from reference manager software
- Internal refinements: user testing, adopting agile software principles