CHORUS and FAIR Data

JOSS / CHORUS Forum: FAIR Data from Funding to Publication
17/18 June 2021

Howard Ratner, Executive Director
hratner@chorusaccess.org
www.chorusaccess.org | @chorusaccess.org
CHORUS 2021 Goals

- Scale OA Compliance
- Make Open Research Work
- Develop Data Metrics
- Improve Metadata Quality
- Connect Stakeholders

PUBLISHERS

INSTITUTIONS

FUNDERS

Advancing Public Access to Research | www.chorusaccess.org
Consuming IDs

- ORCID ID
- Dataset DOIs
- Funder ID
- Content DOI

Monitor and Audit Public Access

- CHORUS
- Crossref
- Publishers
- SCHOLAR
- DataCite
- NSF
- USGS
- NIH
- PORTICO

Coming Soon

Organization IDs

Grant ID

Content DOI

Preservation Status

Funder Portal link
Publication and Compliance Data Services

API, metadata feeds and dashboards for monitoring and tracking open research information.

- Funder Reports
- Institution Reports
- Publisher Reports
- Dataset Reports
## Enhanced Dataset Reporting

<table>
<thead>
<tr>
<th>Article DOI</th>
<th>Journal Title</th>
<th>Agency</th>
<th>Dataset Creators</th>
<th>Dataset DOI</th>
<th>Dataset Type</th>
<th>Dataset Repository Name</th>
<th>Dataset Title</th>
<th>Date Collected at Repository</th>
<th>Funder Name</th>
<th>Award #</th>
<th>Award Title</th>
<th>Rights</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1073/pnas.1604386113</td>
<td>Proceedings of the National Academy of Sciences</td>
<td>National Oceanic and Atmospheric Administration, National Science Foundation</td>
<td>Ganguli, Poulomi; Paprotny, Dominik; Hasan, Mehed; Günther, Andreas; Merz, Bruno</td>
<td>10.5880/gfs.4.4.2019.003</td>
<td>GFZ Data Services</td>
<td>Compound flood drivers for northwestern Europe in high-resolution EURO-CORDEX Simulations</td>
<td>2019</td>
<td>Alexander Von Humboldt Foundation, Germany</td>
<td>D05417008</td>
<td>Humboldt Research Fellowship for early career researchers</td>
<td>CC BY 4.0</td>
<td></td>
</tr>
<tr>
<td>10.1186/s12862-019-1465-5</td>
<td>BMC Evolutionary Biology</td>
<td>Directorate for Biological Sciences, National Oceanic and Atmospheric Administration</td>
<td>Schott, Eric; Lella, Santiago; Tsvetan Bachvaroff; L. Amzel; Vasta, Gerardo</td>
<td>10.6084/m9.figshare.c.4583777.v1</td>
<td>Collection</td>
<td>Figshare</td>
<td>Lacking catalase, a protistan parasite draws on its photosynthetic ancestry to complete an antioxidant repertoire with ascorbate peroxidase</td>
<td>2019</td>
<td></td>
<td></td>
<td>CC BY 4.0</td>
<td></td>
</tr>
<tr>
<td>10.1242/jeb.189233</td>
<td>The Journal of Experimental Biology</td>
<td>Hampden-Sydney College; Harris Award; National Oceanic and Atmospheric Administration; North Slope Borough, Alaska,</td>
<td>Werth, Alexander J.; Rita, Diego; Rosario, Michael V.; Moore, Michael J.; Sformo, Todd L.</td>
<td>10.5061/dryad.73m81p</td>
<td>dataset</td>
<td>Dryad</td>
<td></td>
<td>2018</td>
<td>National Science Foundation</td>
<td>1656691</td>
<td></td>
<td>CC0 1.0 Universal (CC0 1.0) Public Domain Dedication</td>
</tr>
</tbody>
</table>
Convening Stakeholder Discussions

- CHORUS Forum: Making the Future of Open Research Work
- CHORUS Forum: New Connections - Research Data to Content
- CHORUS Forum: Open Access Policies and Compliance in a Global Context
- JOSS / CHORUS Forum: FAIR Data from Funding to Publication
- STM/COS/CHORUS: Towards a US Research Data Framework
- JST: Working together to monitor open science in Japan

Finding Common Problems & Help Define Potential Solutions
Helping to improve metadata

Gap Analysis
- ORCID IDs
- Reuse License Metadata
- Publication Dates

Gap Analysis (in planning)
- ORCID IDs?
- Creators?
- Institutions?
- Funders?
CHORUS Publisher Data Availability Policies Index

Over the last few years, publishers have been making their Data Availability Policies known either at the publisher level or at the journal level. These policies range in their mandate, but most require authors to make all data necessary to replicate their study's findings publicly available without restriction at the time of publication. When specific legal or ethical restrictions prohibit public sharing of a data set, authors must indicate how others may obtain access to the data. Often, when submitting a manuscript to the publisher, authors must provide a Data Availability Statement describing compliance with the data availability policy. If the article is accepted for publication, the Data Availability Statement will be published as part of the article.

CHORUS has created a centralized index of our member publishers' policies with links to the publisher's site. This chart will be updated at least annually.

Please let us know about any updates or requested enhancements by contacting us at info@chorusaccess.org.

<table>
<thead>
<tr>
<th>CHORUS Publisher Member</th>
<th>Data Availability Policy URL</th>
<th>Policy Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAAS/Science</td>
<td><a href="https://www.sciencemag.org/authors/science-journals-editors-policies">https://www.sciencemag.org/authors/science-journals-editors-policies</a></td>
<td>Publisher</td>
</tr>
<tr>
<td>American Association of Physicians in Medicine</td>
<td><a href="https://www.aam.org/rg/policies/details.asp?id=329&amp;type=PP">https://www.aam.org/rg/policies/details.asp?id=329&amp;type=PP</a></td>
<td>Publisher</td>
</tr>
<tr>
<td>American Association of Physics Teachers</td>
<td>TBD</td>
<td>Publisher</td>
</tr>
<tr>
<td>American Astronomical Society</td>
<td><a href="https://journals.aas.org/policies/">https://journals.aas.org/policies/</a></td>
<td>Publisher</td>
</tr>
<tr>
<td>Association for Computing Machinery</td>
<td><a href="https://www.acm.org/acm-policies-and-procedures">https://www.acm.org/acm-policies-and-procedures</a></td>
<td>Publisher</td>
</tr>
<tr>
<td>American College of Physicians</td>
<td><a href="https://annals.org/ajm/pages/authors">https://annals.org/ajm/pages/authors</a></td>
<td>Publisher</td>
</tr>
<tr>
<td>American Chemical Society</td>
<td><a href="https://pubs.acs.org/publish/publishing_policies">https://pubs.acs.org/publish/publishing_policies</a></td>
<td>Publisher</td>
</tr>
<tr>
<td>American Diabetes Association</td>
<td><a href="https://professional.diabetes.org/sites/professional.diabetes.org/files/media/2018_data_resource_sharing_policy.pdf">https://professional.diabetes.org/sites/professional.diabetes.org/files/media/2018_data_resource_sharing_policy.pdf</a></td>
<td>Publisher</td>
</tr>
<tr>
<td>American Geophysical Union</td>
<td><a href="https://www.agu.org/Publish-with-AGU/Publish/Author-Resources/Data-for-Authors">https://www.agu.org/Publish-with-AGU/Publish/Author-Resources/Data-for-Authors</a></td>
<td>Publisher</td>
</tr>
<tr>
<td>American Institute of Aeronautics and Astronauts</td>
<td><a href="https://www.aaiaa.org/publications/books/Publication-Policies">https://www.aaiaa.org/publications/books/Publication-Policies</a></td>
<td>Publisher</td>
</tr>
<tr>
<td>AIP Publishing</td>
<td><a href="https://publishing.aip.org/resources/researchers/open-science/research-data-policy/">https://publishing.aip.org/resources/researchers/open-science/research-data-policy/</a></td>
<td>Publisher</td>
</tr>
<tr>
<td>American Meteorological Society</td>
<td><a href="https://www.ametsoc.org/index.cfm/ams/publications/ethical-guidelines-and-ams-policies/data-archiving-and-citation/">https://www.ametsoc.org/index.cfm/ams/publications/ethical-guidelines-and-ams-policies/data-archiving-and-citation/</a></td>
<td>Publisher</td>
</tr>
<tr>
<td>American Mathematical Society</td>
<td>TBD</td>
<td>Publisher</td>
</tr>
<tr>
<td>American Psychological Association</td>
<td><a href="https://www.apa.org/research/responsible/data">https://www.apa.org/research/responsible/data</a></td>
<td>Publisher</td>
</tr>
<tr>
<td>American Physical Society</td>
<td>TBD</td>
<td>Publisher</td>
</tr>
<tr>
<td>American Society of Mechanical Engineers</td>
<td><a href="https://www.asme.org/publications-submissions/journals/information-for-authors/data-policy">https://www.asme.org/publications-submissions/journals/information-for-authors/data-policy</a></td>
<td>Publisher</td>
</tr>
</tbody>
</table>

https://www.chorusaccess.org/resources/chorus-for-publishers/publisher-data-availability-policies-index/
Accelerating Open and FAIR Data Practices Across the Earth, Space, and Environmental Sciences: A Pilot with the NSF to Support Public Access to Research Data

Two-year project aimed at implementing FAIR data practices across the Earth, space, and environmental sciences such that, by the end of the project:

- **Scope**: AGU data citations for research funded by NSF grants are captured in the NSF Public Access Repository (PAR 2.0)
- **Knowledge of leading practices and workflows around data citation are well known across the AGU community**
- **Ensure NSF earth science data are publicly accessible, linked to publications and grants, and credited**

• Present dataset reports:
  - linked articles,
  - name and type of dataset,
  - repository and creator names,
  - subject classification, and
  - reuse license information

• Help AGU and others improve data citation deposit into the NSF Public Access Repository (NSF PAR)
• Develop and further open data citation best practices

**Participants:**
AGU | NSF | CHORUS | Dryad | Earth Science Information Partners (ESIP) | Wiley

All learnings and practices from the work will be openly shared
Apply machine-learning and natural language processing techniques that searches publications to:

- Find what datasets are in the publications
- Show how they have been used
- Find other experts who have used the data
- Identify other related datasets
- Show impact of funded datasets

https://www.kaggle.com/c/coleridgeinitiative-show-us-the-data/
How might we?

Project Goal

How might we enable government agencies to show how, why and by whom their data are being used?

The objective + ask

Demonstrate the value of data as a strategic asset to the public and legislators.

Design a data scorecard that can serve as a source of information about users, data sources and applications of data.

Credit: Julia Lane, Coleridge Initiative
Improve model through AI and community

Use Machine Learning to identify links
- Datasets
- Publications

Experts validate links
- Initial RCKG

Community uses and improves
- Enhanced RCKG

Calculate measures and topics
- Improve algorithms

Credit: Julia Lane, Coleridge Initiative
Approach: Deploy new technology?

Product Market Competition and Human Resource Practices in the Retail Food Sector

ELIZABETH DAVIS, MATTHEW FREEDMAN, JULIA LANE, BRIAN MCCALL, NICOLE NESTORIAR, and TIMOTHY PARK*

In the wake of Wal-Mart and other mass merchandisers’ entry into food retailing, the nature of competition in the industry has changed radically. Using longitudinal data on workers and firms to construct measures of compensation and churning for traditional food retailers, this paper examines how these measures change in response to mass merchandiser entry. While there is considerable heterogeneity across retail food establishments, human resource practices are persistent even in the face of new external competition.

Introduction

Whereas Wal-Mart employees start at the same salary as unionized employees in similar lines of work, they make 25 percent less than their unionized counterparts

Data and Measurement

Constructing a data set that permits the analysis of the effects of changes in the product market on firm specific compensation policies requires information on firms, their workers, and the product markets in which they

* This is different from the approach used by Neumark et al. (2005) since the research question is broader than examining the impact of the entry of one particular firm.

Credit: Julia Lane, Coleridge Initiative
What does such a search generate?

- Dataset name (and aliases)
- Research topics
- Authors (and associated metadata)
- Related datasets
- Publication name (and associated metadata)

*Credit: Julia Lane, Coleridge Initiative*
Usage Indicators for
USDA Agricultural Resource Management Survey

<table>
<thead>
<tr>
<th>TOPICS</th>
<th>17</th>
<th>AUTHORS</th>
<th>539</th>
<th>JOURNALS</th>
<th>86</th>
<th>WITH OTHER DATASETS</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agricultural Research...</td>
<td>73</td>
<td>Ashok K. Mishra</td>
<td>13</td>
<td>SSRN Electronic Journ...</td>
<td>48</td>
<td>Census of Agricultur...</td>
<td>41</td>
</tr>
<tr>
<td>Farm Household Well-...</td>
<td>54</td>
<td>William D. McBride</td>
<td>13</td>
<td>Journal of Agricultu...</td>
<td>27</td>
<td>Rural-Urban Continuu...</td>
<td>4</td>
</tr>
<tr>
<td>Organic Agriculture</td>
<td>34</td>
<td>Nigel D. Key</td>
<td>12</td>
<td>Agricultural and Res...</td>
<td>22</td>
<td>Survey of Industrial...</td>
<td>1</td>
</tr>
<tr>
<td>Poverty &amp; Income Vol...</td>
<td>16</td>
<td>Hisham S. El-Osta</td>
<td>11</td>
<td>Agricultural Finance...</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Climate Change</td>
<td>9</td>
<td>James M. MacDonald</td>
<td>8</td>
<td>Sustainability</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Food Service Industry...</td>
<td>8</td>
<td>Ani L. Katchova</td>
<td>8</td>
<td>American Journal of ...</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Credit: Julia Lane, Coleridge Initiative
Join Our Community!
Contact us today

Howard Ratner, Executive Director
hratner@chorusaccess.org

Mark Robertson, Director Asia-Pacific Development
mrobertson@chorusaccess.org

www.chorusaccess.org | @chorusaccess.org