

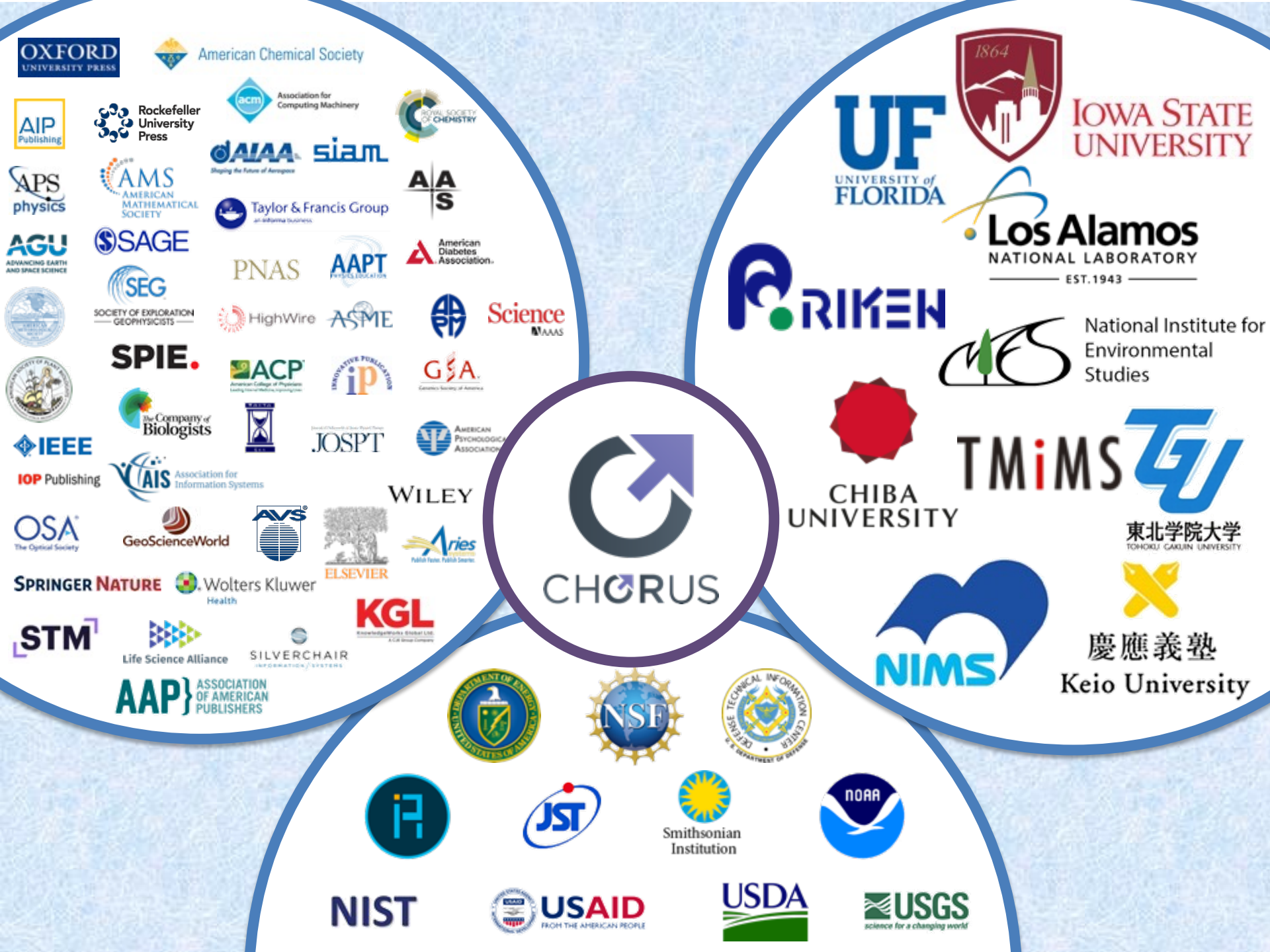


CHORUS and FAIR Data

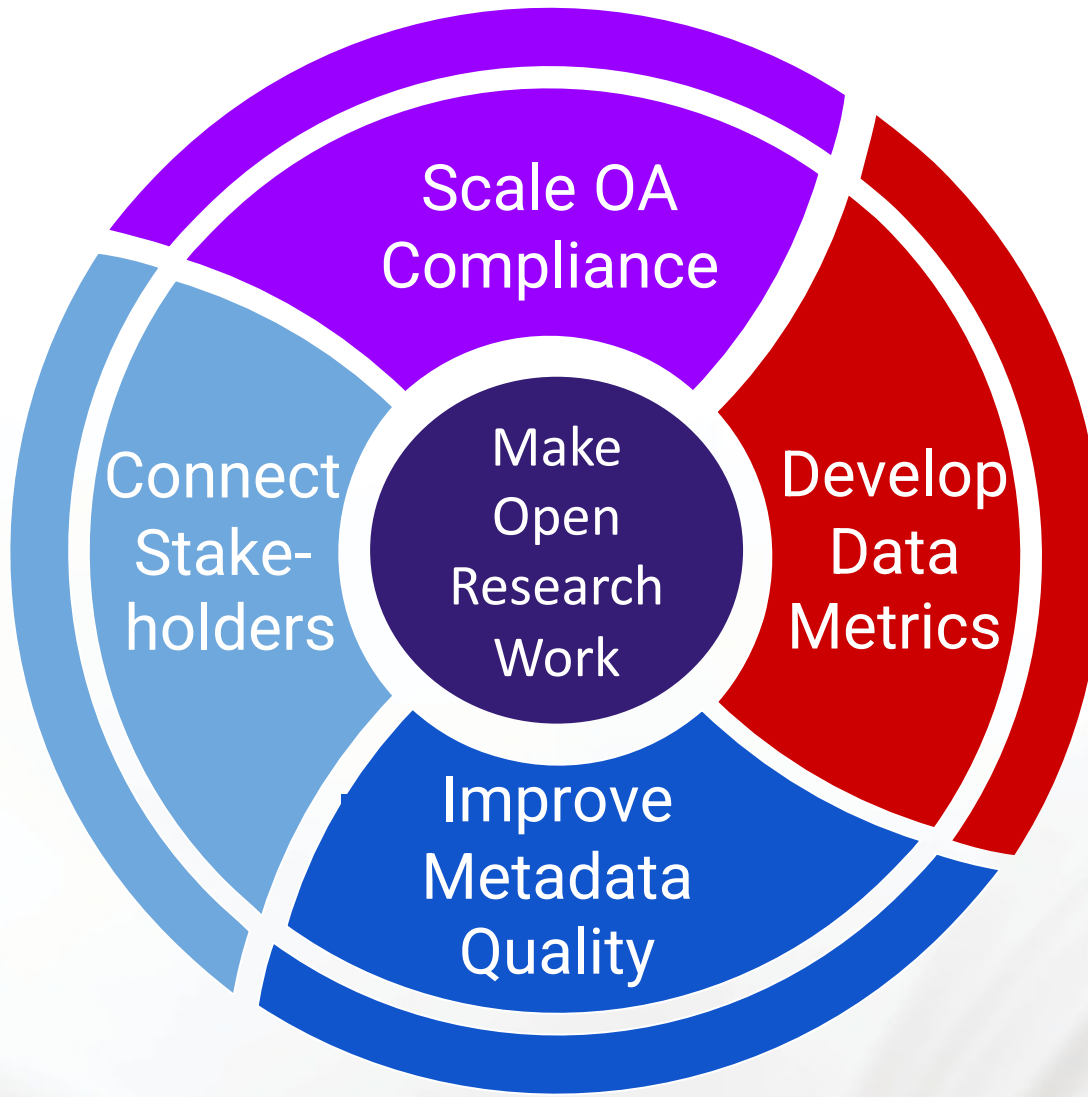
JOSS / CHORUS Forum: FAIR Data from Funding to Publication
17/18 June 2021

Howard Ratner, Executive Director
hratner@chorusaccess.org
www.chorusaccess.org | [@chorusaccess.org](https://twitter.com/chorusaccess.org)

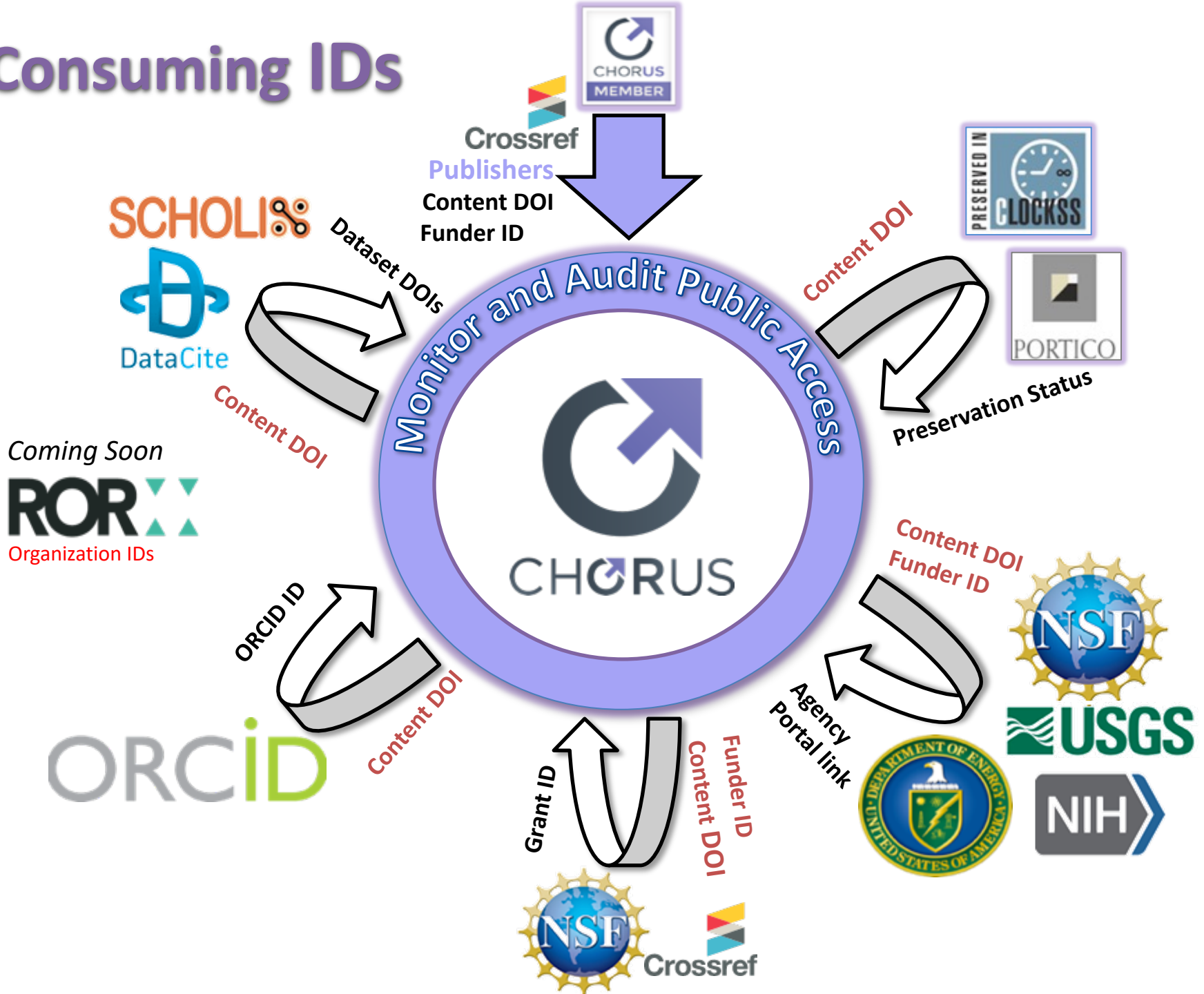




CHORUS 2021 Goals



Consuming IDs



Publication and Compliance Data Services

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

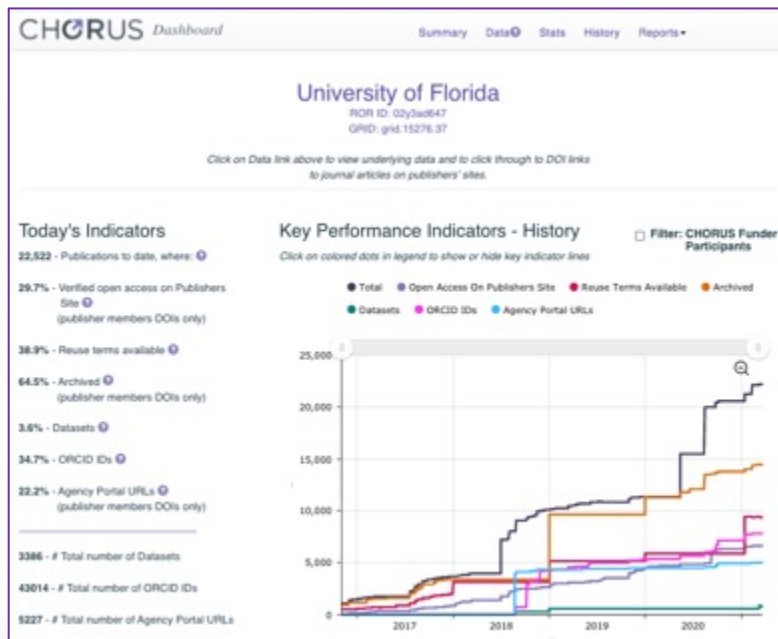


Funder Reports

Institution Reports

Publisher Reports

Dataset Reports



University of Florida

Record level information about Publisher Member content.
Click on DOIs to be taken to articles on publisher site.

Indicator History Details

3/25/2021

VERIFIED OPEN ACCESS ON PUBLISHERS SITE - YES



Search:

Updated	Grant ID	Archived	Datasets (beta)
journal-article	3/7/2021		
	NIH:U54GM111274 NIH:R01 AI102939-05	Yes	http://dx.doi.org/10.7927/h4cc0xnrv
	NSF:EAR1141840 NSF:OCE1058858 NSF:1141840 NSF:1058858	Yes	http://dx.doi.org/10.7288/v4/magic/16749 http://dx.doi.org/10.7288/v4/magic/12152 http://dx.doi.org/10.7288/v4/magic/16760 http://dx.doi.org/10.7288/v4/magic/16761
	NSF:(EAR0608998 and EAR0609339)	Yes	http://dx.doi.org/10.7288/v4/magic/16345 http://dx.doi.org/10.7288/v4/magic/16347 http://dx.doi.org/10.7288/v4/magic/16346

EARTH-DATA Other DAACS

NASA SOCIOECONOMIC DATA AND APPLICATIONS CENTER (SEDAC)
A Data Center in NASA's Earth Observing System Data and Information System (EOSDIS) — Hosted by CIRES at Columbia University

DATA MAPS THEMES RESOURCES SOCIAL MEDIA ABOUT HELP

Population Dynamics

Collection Overview

Data Sets (7)

Global Population Count Grid Time Series Estimates, v1 (1970–2000)

Set Overview Data Download Maps Map Services Documentation Metadata

Purpose:
To provide back-cast population count estimates at 30 arc-second (~1 km) resolution.

Abstract:
The Global Population Count Grid Time Series Estimates provide a back-cast time series of population grids based on the year 2000 population grid from SEDAC's Global Rural-Urban Mapping Project, Version 1 (GRUMPv1) data set. The grids were created by using rates of population change between decades from the coarser resolution History Database of the Global Environment (HYDE) database to back-cast the GRUMPv1 population count grids. Mismatches between the spatial extent of the HYDE calculated rates and GRUMPv1 population data were resolved via infilling rate cells based on a focal mean of values. Finally, the grids were adjusted so that the population totals for each country equal the UN World Population Prospects (2008 Revision) estimates for that country for the respective year (1970, 1980, 1990, 2000). These data do not represent census observations for the years prior to 2000, and therefore can at best be thought of as estimations of the populations in given locations. The population grids are consistent internally within the time series, but are not recommended for use in creating longer time series with any other population grids, including GRUMPv1, Gridded Population of the World, Version 4 (GPWv4), or map-SEDAC developed population grids. These population grids served as an input to SEDAC's Global Estimated Net Migration Grids by Decade: 1970-2000 data set.

2 of 4

EarthRef.org MagIC GERM SBN FeMO SCC ERESE ERDA References Users Log In / Register

Magnetism Information Consortium (MagIC)
Promoting information technology infrastructures for the international paleomagnetic, geomagnetic and rock magnetic community.

Home About Technology Grand Challenges Workshops Links Report an Issue on GitHub Help Contact

Contributions 1 Locations 1 Sites 1,807 Samples 2,404 Specimens 2,404 Experiments 2,415 Measurements 4,007

Search MagIC ID: 16760 Search Clear Download Results

Filters Clear Filters

Summaries 1 Map 1 Recently Contributed First

Publication Year: Anytime to 2021

Author

Contributor

Geospatial: Lat: -90 to 90 deg Lon: -180 to 180 deg

Age: 0 Ma to 200 Ma

Absolute Paleointensity

Method Code

External Database

Location Type

Download MagIC Contribution Link EarthRef Data DOI Link Version Data Model Date Co

Download earthref.org/MagIC/16761 10.7288/v4/magic/16761 2 3.0 February 23, 2020 US

Download earthref.org/MagIC/16760 10.7288/v4/magic/16760 1 3.0 February 23, 2020 US

Geologic: Continental Shelf Class: Sedimentary Age: 0 - 13 Ma No Intensity Data Method Codes: GM-PMAG-POL, GM-FOSSIL, SO-V, PS-C, DRILL-ICDP, SP-SS-C, AL-H, LP-AN-MG, SO-MG, LP-X, SE-...

Geographic: Indian Ocean, Wilkes Land Marg... Lithology: Silicate Clay, Silty...

Citations: 10.1029/2019PA002309

L. Tauxe, S. Sugliani, F. Jiménez-Espejo, C. Escutia, C.P. Cook, T. van de Flierdt, M. Ival (2019), Geology of the Wilkes land sub-basin and stability of the East Antarctic ice Sheet: Insights from rock magnetism at ICDP Site U1361. Earth and Planetary Science Letters 412:61-69. doi:10.1016/j.epsl.2014.12.034

Crossref Citation Count: 7

Sponsored by NSF. Updated on Apr 9, 2021. Having trouble? Email Us Powered by FIESTA Unless otherwise noted, EarthRef.org content is licensed under CC BY 4.0

Enhanced Dataset Reporting

NEW

Article DOI	Journal Title	Agency	Dataset Creators	Dataset DOI	Dataset Type	Dataset Repository Name	Dataset Title	Date Collected at Repository	Funder Name	Award #	Award Title	Rights
10.1073/pnas.1604386113	Proceedings of the National Academy of Sciences	National Oceanic and Atmospheric Administration, National Science Foundation	Ganguli, Poulomi; Paprotny, Dominik; Hasan, Mehedi; Güntner, Andreas; Merz, Bruno	10.5880/gfz.4.4.2019.003		GFZ Data Services	Compound flood drivers for northwestern Europe in high-resolution EURO-CORDEX Simulations	2019	Alexander von Humboldt Foundation, Germany	D05417008	Humboldt Research Fellowship for early career researchers	CC BY 4.0
10.1186/s12862-019-1465-5	BMC Evolutionary Biology	Directorate for Biological Sciences, National Oceanic and Atmospheric Administration	Schott, Eric; Lella, Santiago; Tsvetan Bachvaroff; L. Amzel; Vasta, Gerardo	10.6084/m9.figshare.c.4583777.v1	Collection	Figshare	Lacking catalase, a protistan parasite draws on its photosynthetic ancestry to complete an antioxidant repertoire with ascorbate peroxidase	2019				CC BY 4.0
10.1242/jeb.189233	The Journal of Experimental Biology	Hampden-Sydney College; Harris Award; National Oceanic and Atmospheric Administration; North Slope Borough, Alaska,	Werth, Alexander J.; Rita, Diego; Rosario, Michael V.; Moore, Michael J.; Sformo, Todd L.	10.5061/dryad.73rm81p	dataset	Dryad		2018	National Science Foundation	1656691		CC0 1.0 Universal (CC0 1.0) Public Domain Dedication

Convening Stakeholder Discussions

CHORUS Forum:
Making the Future
of Open Research
Work

JOSS / CHORUS
Forum: FAIR Data
from Funding to
Publication

CHORUS Forum:
New Connections
- Research Data to
Content

Finding
Common
Problems

Help Define
Potential
Solutions

JST: Working
together to
monitor open
science in Japan

CHORUS Forum:
Open Access Policies
and Compliance in
a Global Context

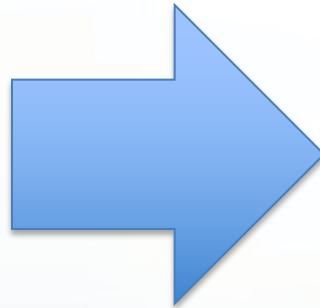
STM/COS/CHORUS:
Towards a US Research
Data Framework

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

NEW!

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.



CHORUS Publisher Member	Data Availability Policy URL	Policy Type
AAAS/Science	https://www.sciencemag.org/authors/science-journals-editorial-policies	Publisher
American Association of Physicists in Medicine	https://www.aapm.org/org/policies/details.asp?id=329&type=PP	Publisher
American Association of Physics Teachers	TBD	
American Astronomical Society	https://journals.aas.org/policies/	Publisher
Association for Computing Machinery	https://www.acm.org/acm-policies-and-procedures	Publisher
American College of Physicians	https://annals.org/aim/pages/authors	Publisher
American Chemical Society	https://publish.acs.org/publish/publishing_policies https://publish.acs.org/publish/author_guidelines?coden=acscii#data_requirements	Publisher
American Diabetes Association	https://professional.diabetes.org/sites/professional.diabetes.org/files/media/2018_data_resource_sharing_policy.pdf	Publisher
American Geophysical Union	https://www.agu.org/Publish-with-AGU/Publish/Author-Resources/Data-for-Authors	Publisher
American Institute of Aeronautics and Astronautics	https://www.aiaa.org/publications/books/Publication-Policies	Publisher
AIP Publishing	https://publishing.aip.org/resources/researchers/open-science/research-data-policy/	Publisher
American Meteorological Society	https://www.ametsoc.org/index.cfm/ams/publications/ethical-guidelines-and-ams-policies/data-archiving-and-citation/	Publisher
American Mathematical Society	TBD	
American Psychological Association	https://www.apa.org/research/responsible/data	Publisher
American Physical Society	TBD	
American Society of Mechanical Engineers	https://www.asme.org/publications-submissions/journals/information-for-authors/data-policy	Publisher

<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



Grant 2025364

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

<https://fromtheprow.agu.org/nsf-grant-takes-agus-data-and-publishing-practices-to-the-next-level/>

COLERIDGE INITIATIVE

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

Rich Context Project

The screenshot shows the top of a Kaggle competition page for the 'Coleridge Initiative - Show US the Data'. The header includes the competition title, a subtitle 'Discover how data is used for the public good', and a prize money of '\$90,000'. It also mentions '1,321 teams' and '19 days to go (12 days to go until merger deadline)'. Below the header is a navigation bar with links for 'Overview', 'Data', 'Code', 'Discussion', 'Leaderboard', and 'Rules', and a 'Join Competition' button. The main content area is titled 'Overview' and contains a table with the following rows: 'Description', 'Evaluation', 'Timeline', 'Prizes', 'Code Requirements', and 'Coleridge Initiative'. The 'Description' row is highlighted, and its content is displayed in the main text area. The content describes the competition's goal: to challenge data scientists to show how publicly funded data are used to serve science and society. It mentions that evidence through data is critical for addressing threats like pandemics, climate change, Alzheimer's disease, child hunger, and increasing food production. It also notes that much of the information about data necessary to inform evidence and science is locked inside publications. The text further explains that the competition will build an open and transparent approach, showing how public data are being used in science and helping the government make wiser, more transparent public investments. It mentions that the results will help move researchers and governments from using ad-hoc methods to automated ways of finding out what datasets are being used to solve problems, what measures are being generated, and which researchers are the experts. Previous competitions have shown that it is possible to develop algorithms to automate the search and discovery of references to data. Now, the goal is to have the Kaggle community develop the best approaches to identify critical datasets used in scientific publications.

Overview	
Description	This competition challenges data scientists to show how publicly funded data are used to serve science and society. Evidence through data is critical if government is to address the many threats facing society, including; pandemics, climate change, Alzheimer's disease, child hunger, increasing food production, maintaining biodiversity, and addressing many other challenges. Yet much of the information about data necessary to inform evidence and science is locked inside publications.
Evaluation	
Timeline	
Prizes	Can natural language processing find the hidden-in-plain-sight data citations? Can machine learning find the link between the words used in research articles and the data referenced in the article?
Code Requirements	
Coleridge Initiative	Now is the time for data scientists to help restore trust in data and evidence. In the United States, federal agencies are now mandated to show how their data are being used. The new Foundations of Evidence-based Policymaking Act requires agencies to modernize their data management. New Presidential Executive Orders are pushing government agencies to make evidence-based decisions based on the best available data and science. And the government is working to respond in an open and transparent way . This competition will build just such an open and transparent approach. The results will show how public data are being used in science and help the government make wiser, more transparent public investments. It will help move researchers and governments from using ad-hoc methods to automated ways of finding out what datasets are being used to solve problems, what measures are being generated, and which researchers are the experts. Previous competitions have shown that it is possible to develop algorithms to automate the search and discovery of references to data. Now, we want the Kaggle community to develop the best approaches to identify critical datasets used in scientific publications.

<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?

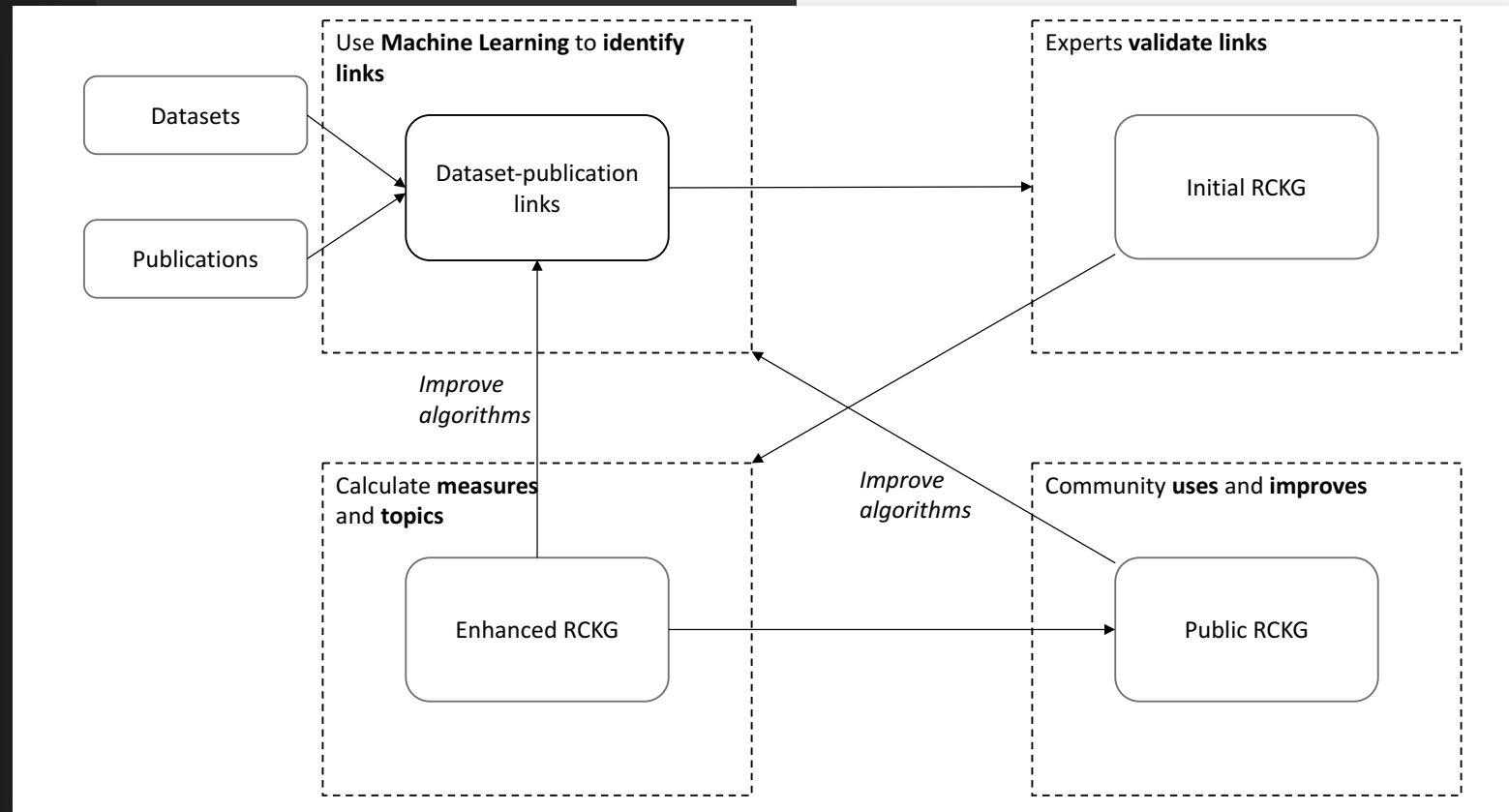
The objective + ask

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



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 NESTORIAK,
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.

358 / ELIZABETH DAVIS ET AL.

operate. We rely on a new linked employer-employee database that also provides information on the location and industry of firms: the Longitudinal Employer-Household Dynamics (LEHD) Program at the U.S. Census Bureau (described in detail by Abowd et al. (2005) and Haltiwanger, Lane, and Spletzer (2006)). Briefly, the LEHD data consist of quarterly records of the employment and earnings of all workers who are covered by U.S. state unemployment insurance (UI) systems in the 1990s and early 2000s.⁵ About 96 percent of private wage and salary employment is covered by these

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

Used by 232 Publications based on verified attributions for v.2, updated 01-02-2021

 **17**
Topics

 **539**
Authors

 **86**
Journals

 **3**
With Other Datasets

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

Proof of Concept

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](https://twitter.com/chorusaccess)

