

Monitoring policy compliance with Wikidata and Scholia

Lane Rasberry & Daniel Mietchen
25 October 2021

[DOI: 10.5281/zenodo.5594663](https://doi.org/10.5281/zenodo.5594663)





Lane Rasberry
Wikimedian in Residence
School of Data Science
University of Virginia



Daniel Mietchen
Researcher on Open Science
Ronin Institute

Structure of the talk

1. assessing policy compliance
2. Scholia's unique position
3. development directions

Assessing policy compliance



THEORY

Problem:

Lack of **FAIR & open data** about

- funded projects
 - project outcomes
 - applicable policies
 - components of compliance
- hinders compliance assessments.

Solution:

- Use a **FAIR & open data** platform that can handle
 - funded projects
 - project outcomes
 - applicable policies
 - components of compliance
- Aggregate and visualize relevant data

Solution:

- Use a **FAIR & open data** platform that can handle
 - funded projects
 - project outcomes
 - applicable policies
 - components of compliance
- Aggregate and visualize relevant data



Assessing policy compliance

PRACTICE



Example funding record for a project



Quantifying the Role of
Tropical Forests in the Global
Carbon Balance and Future
Climate Change

Date Awarded: Nov 2007

Amount: \$3,893,432

Term: 48 months

Grant ID: GBMF1656

Funding Area: Environmental Conservation

(paragraph description of the project)

CC BY-NC-ND 4.0

<https://www.moore.org/grant-detail?grantId=GBMF1656>

Wikidata entry
on the
project



<u>start time</u>	November 2007
<u>funder</u>	Moore Foundation
<u>duration</u>	48 months
<u>subject</u>	environmental protection
<u>part of</u>	Amazon Forest Inventory
<u>sponsor</u>	University of Leeds

Wikidata entry
on the
project's paper



author

Gabriela Lopez-Gonzalez

license

CC BY-NC 4.0

funder

(none mentioned)

DOI

(none)

Language

Spanish

Which policies apply?

Institutions

Funders

Disciplines

Journals

Countries

The foundation requires that a final (post-print²) version of all peer-reviewed articles produced as a result of research supported, either in entirety or in part, by the foundation's funding, be made publicly and freely available (open access¹, or OA) within 12 months of publication.

<https://www.moore.org/docs/default-source/Grantee-Resources/open-access-policy-2017.pdf>

1 and 2 are defined in the PDF (for humans, not machines)




Wikidata entry on the funder's open access policy

Wikidata




(Q109204635) ☆ Watch

English


Gordon and Betty Moore Foundation
Open Access Policy
open-access policy 

∨ In more languages

Statements

instance of  open access policy 0 references	signatory  Gordon and Betty Moore Foundation 0 references	full work available at URL  https://www.moore.org/docs/default-source/Grantee-Resources/open-access-policy-2017.pdf 0 references
---	--	---

Identifiers

Sherpa Juliet ID  246 0 references

Is CC BY-NC 4.0 compliant with GBMF's OA policy?

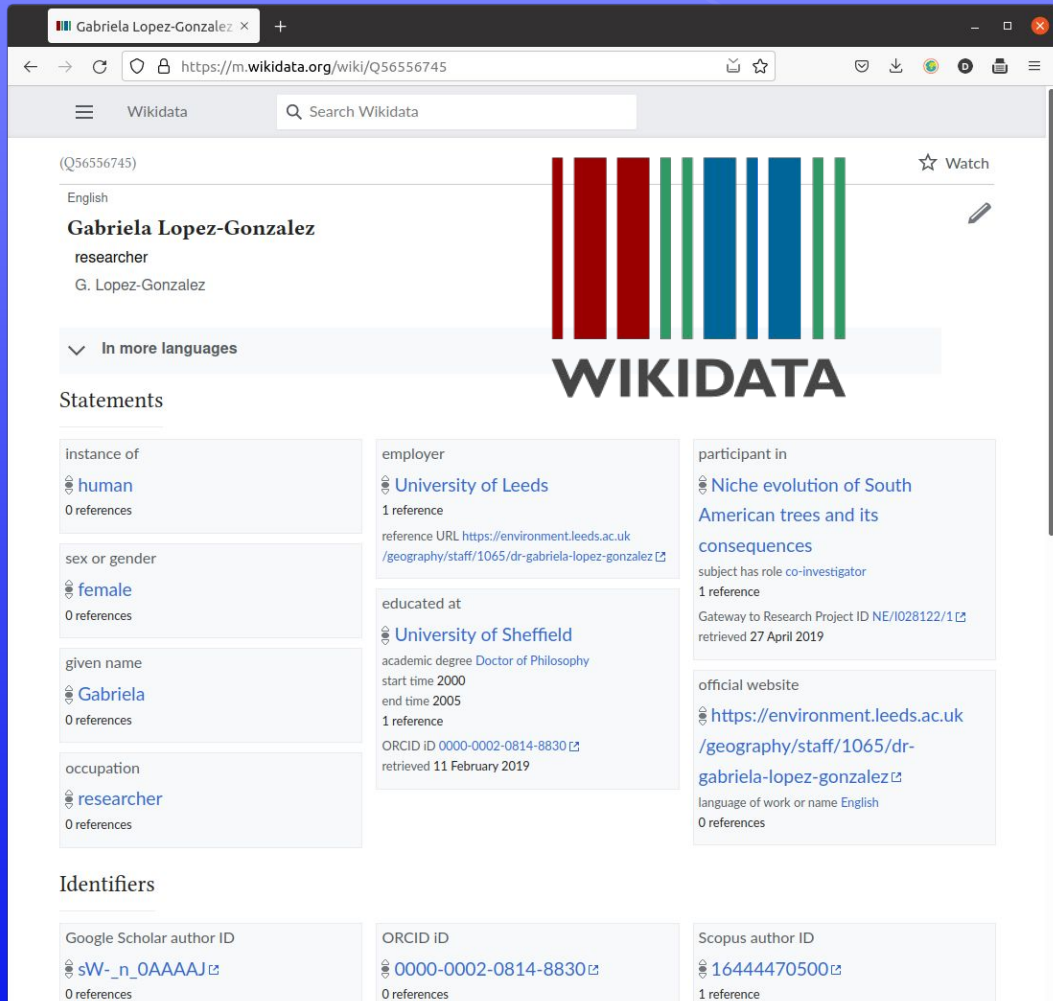
*grantees are expected to provide the **widest possible access** to their published research*

*Open access (OA) content is digital, online, free of charge, and **free of most copyright and licensing restrictions**.*

<https://www.moore.org/docs/default-source/Grantee-Resources/open-access-policy-2017.pdf>

How to make **such** criteria machine actionable?

Wikidata entry on the author



The screenshot shows a web browser window displaying the Wikidata entry for Gabriela Lopez-Gonzalez. The browser's address bar shows the URL <https://m.wikidata.org/wiki/Q56556745>. The Wikidata interface includes a search bar, a 'Watch' button, and a 'WIKIDATA' logo. The entry is in English and identifies Gabriela Lopez-Gonzalez as a researcher. It lists several statements: instance of human, sex or gender female, given name Gabriela, occupation researcher, employer University of Leeds, participant in Niche evolution of South American trees and its consequences, educated at University of Sheffield, and official website <https://environment.leeds.ac.uk/geography/staff/1065/dr-gabriela-lopez-gonzalez>. Identifiers include Google Scholar author ID [sW-_n_0AAAAJ](#), ORCID ID [0000-0002-0814-8830](#), and Scopus author ID [16444470500](#).

Gabriela Lopez-Gonzalez (Q56556745)

English

Gabriela Lopez-Gonzalez
researcher
G. Lopez-Gonzalez

In more languages

Statements

instance of
[human](#)
0 references

sex or gender
[female](#)
0 references

given name
[Gabriela](#)
0 references

occupation
[researcher](#)
0 references

employer
[University of Leeds](#)
1 reference
reference URL <https://environment.leeds.ac.uk/geography/staff/1065/dr-gabriela-lopez-gonzalez>

educated at
[University of Sheffield](#)
academic degree Doctor of Philosophy
start time 2000
end time 2005
1 reference
ORCID iD [0000-0002-0814-8830](#)
retrieved 11 February 2019

participant in
[Niche evolution of South American trees and its consequences](#)
subject has role co-investigator
1 reference
Gateway to Research Project ID [NE/I028122/1](#)
retrieved 27 April 2019

official website
<https://environment.leeds.ac.uk/geography/staff/1065/dr-gabriela-lopez-gonzalez>
language of work or name English
0 references

Identifiers

Google Scholar author ID
[sW-_n_0AAAAJ](#)
0 references

ORCID ID
[0000-0002-0814-8830](#)
0 references

Scopus author ID
[16444470500](#)
1 reference

Scholia profile for the author

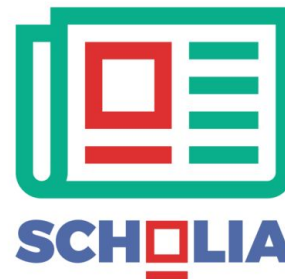
author

Gabriela Lopez-Gonzalez (Q56556745)

 <https://orcid.org/0000-0002-0814-8830>

Table of Contents

- List of publications
 - Number of publications per year
 - Number of pages per year
- Topics
 - Topic scores
 - Topics of authored works
 - Topics-works matrix
- Venue statistics
- Review statistics
- Co-author graph
- Co-author map
- Other locations
- Timeline
- Academic tree
- Citation statistics
 - Most cited works
 - Citations by year
 - Citing authors
- Associated images
- Events



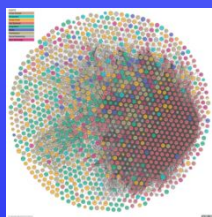
<https://scholia.toolforge.org/author/Q56556745>

Linked Open Data in action!

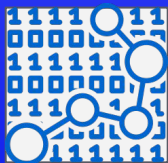
What is Scholia?



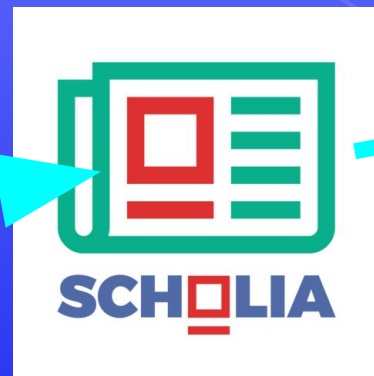
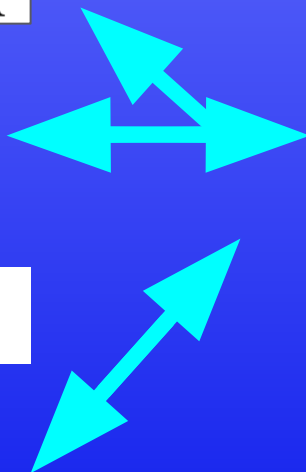
WIKIPEDIA



LOD cloud

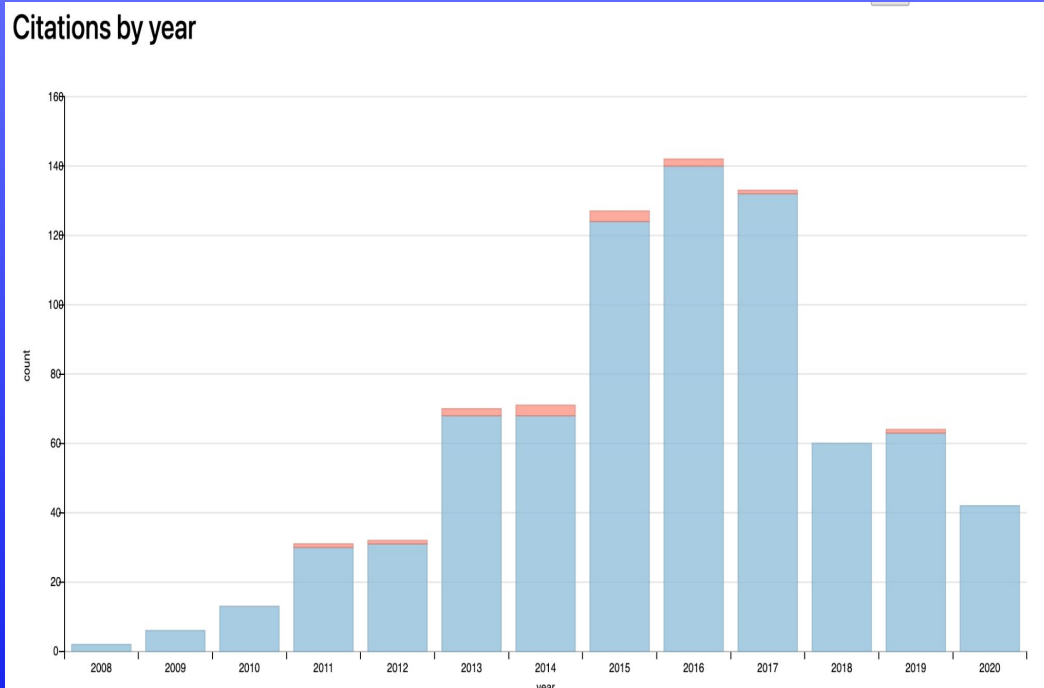


other data



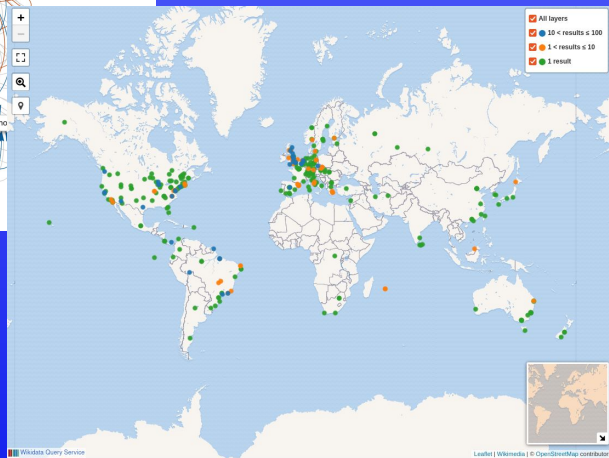
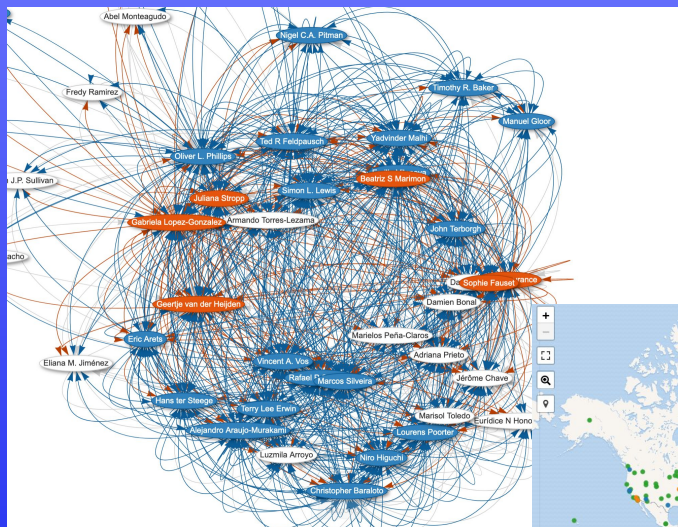
profile

Author's publication history



- Based on FAIR & open data about
 - Author
 - Papers
 - Citations
 - Publication dates
- ~ complete only for pilot use cases

Co-author network



Based on FAIR & open data about

- Author
- Papers
- Co-authors
- Genders
- Affiliations
- Geolocations

~ complete only for pilot use cases

<https://scholia.toolforge.org/author/Q56556745#coauthors-header>

<https://scholia.toolforge.org/author/Q56556745#coauthor-map-header>

Some pilot use cases

- Zika virus
- COVID-19
- Invasion biology
- Nobel Prize winners
- Vanderbilt University
- Scholia example pages
- WikiProject Women in Red
- Scholia profiles linked from Wikipedia
- ORFG-funded research?

Observations with Scholia



- paper available online
- paper has DOI
- paper indexed in [database]
- paper has copyright license
- project has website
- author has ORCID
- author has demographic data



- paper metadata not available
- copyright license non-standard
- paper fails to mention funder
- paper / project / PI not associated with grant catalog entry



Possible Scholia Queries

count of research projects sponsored by ORFG members

<https://w.wiki/4GfX>

Recent citations to publications funded by ORFG members

<https://w.wiki/4GfK>

Genders of authors of ORFG-funded papers

<https://w.wiki/4GfP>


**Scholia's
unique
position**





Scholia is a **FAIR & open**
scholarly profiling service
built around the **popular**
Wikipedia platform

Comparison of scholarly profile tools



Google
Scholar



SEMANTIC SCHOLAR



Scopus®



WEB OF SCIENCE™

Scholia special because...

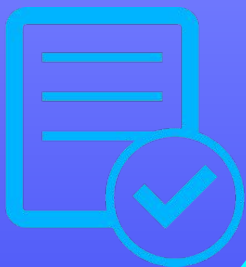
- not a silo; transparent
- FAIR and open
- Wikipedia brand & culture
- high community engagement
- most multidisciplinary
- can help address biases
- can innovate with ethical issues
- multilingual

Scholia on policy compliance

- ⬡ Left: Profile for a **topic**
 - Policy compliance
 - <https://scholia.toolforge.org/topic/Q58028485>

- ⬡ Right: Profile for an individual **work**
 - Breaking the rules: understanding non-compliance with policies and guidelines
 - <https://scholia.toolforge.org/work/Q58491683>





to check
compliance



make data
FAIR



to partner
with Wiki



make it open
too

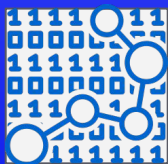
What is Scholia?



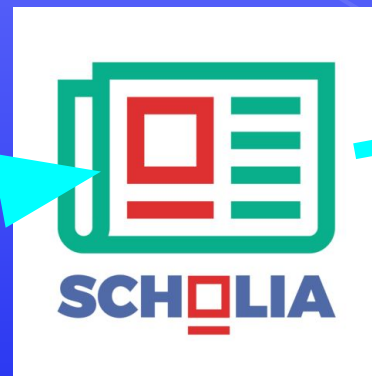
WIKIPEDIA



LOD cloud



other data



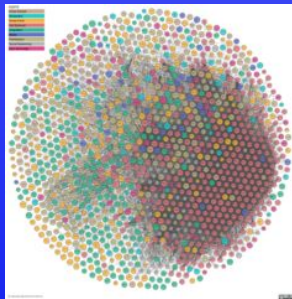
profile

FAIR and open ORFG data

Content development

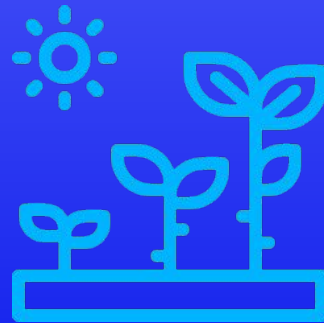
Ready for import

- citation metadata
- grants metadata
- policy metadata



Needs curation

- author disambiguation
- topic discovery
- language translation



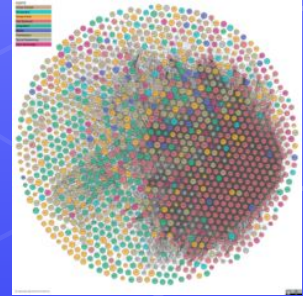
Needs discussion

- demographics
- unfavorable findings
- ontology refinement



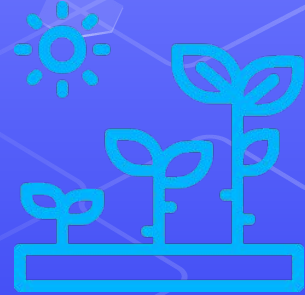
Ready to import data

- consistent data model
 - ideally mapped to external standard
- scholarly literature citation metadata
 - author - title - topic
- list of grant funded projects
 - by funder, by grant recipient
- new media data
 - software, datasets, algorithms



Needs curation

- wiki curation workflows
 - author disambiguation
- machine learning for document sorting
 - topic discovery
- language translation
 - community partnerships



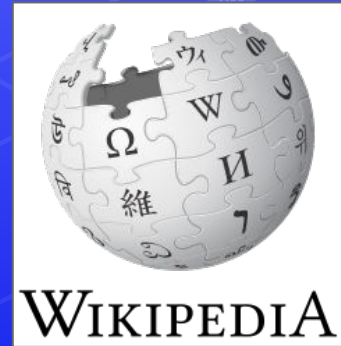
More ethical discourse

Global issues

- Who do we include in discussions?
- Who benefits first?

Big but topical issues

- reporting demographics
- addressing data gaps
- reporting problematic observations



Possible development directions



Possible developments

- Content development
 - Improve content for existing use cases
 - “Model, import, curate, discuss” for further use cases
- Community development
 - Support curators of compliance-related information in curating collaboratively
 - Promoting Wikidata/ Scholia as a FAIR & open approach to scientometrics
- Policy development
 - Make policies more FAIR, e.g. by standardizing key policy elements
 - Make policies and infrastructure more aware of each other
- Platform development
 - Automation and scaling, support for reproducible snapshots
 - Dedicated profiles for policies, with compliance stats

Acknowledgments

- ⬡ Wikimedia community
- ⬡ [Scholia team](#)
- ⬡ Alfred P. Sloan Foundation ([G-2019-11458](#))

Further reading

- ⬡ [Wikidata:Scholia](#)
- ⬡ [Scholia about Scholia](#)
- ⬡ [Scholia on GitHub](#)

Thanks!

Any questions?

You can find us at:

[@LaneRaspberry](#)

lr2ua@virginia.edu

[@EvoMRI](#)

daniel.mietchen@ronininstitute.org

[DOI: 10.5281/zenodo.5594663](https://doi.org/10.5281/zenodo.5594663)

Bonus slides

They follow below to facilitate discussion



Possible development directions



FAIR policies

Policies that are

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable

Example: Policy elements of OA policies

Policy elements

- Who or what is regulated?
- What is the policy based on?
- Are there stipulations about
 - Licensing
 - Embargoes
 - Data / Software
 - Versioning
 - ... ?
 - If so, what is stipulated?

FAIRification issues

- Can this be modularized?
- Can this be expressed in a standard fashion?
 - Controlled vocabularies
 - Ontologies
 - Persistent identifiers
- Can compliance of a policy with a standard be assessed?

Demographic data

**Wiki observations on
diversity, equity, and inclusion**

Wikidata properties for a researcher



WIKIPEDIA



expected

ORCID

occupation / field of work

educated at

affiliation / employer

country of citizenship

languages spoken

perhaps surprising

personal pronoun

sex or gender

place of birth

ethnic group

religion

sexual orientation

Demographic data: ethical issues

Good things

- supports diversity, equity, and inclusion
- credits / showcases underserved communities
- promotes collaboration, humility, transparency

Bad or odd things

- Is gender / ethnicity private?
- cannot come from institutions
- structured data can have weird consequences

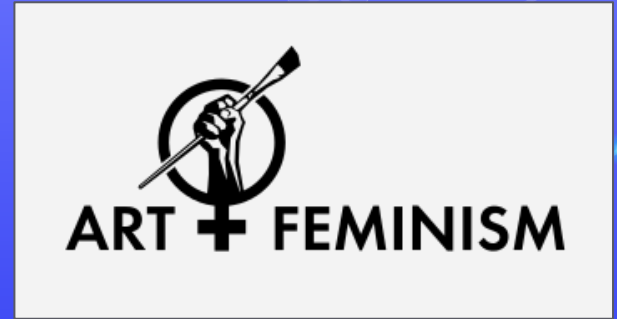
Sources of demographic data



WIKIPEDIA



Women in Red



Sources of demographic data



COMMENTARY | VOLUME 184, ISSUE 3, P561-565, FEBRUARY 04, 2021



Other queries



How Scholia can profile the relationship between projects and outcomes

Potential queries (i)

- Recent publications with funding acknowledgment to ORFG members: <https://w.wiki/4GfG>
- Recent citations to publications funded by ORFG members <https://w.wiki/4GfK>
- Genders of authors of ORFG-funded papers <https://w.wiki/4GfP>
- Number of items with sponsors: total <https://w.wiki/4GfZ> , total ranked list <https://w.wiki/4Gfb> , ORFG <https://w.wiki/4GfY> , ORFG ranked list <https://w.wiki/4GfX>
- Solely funders: <https://w.wiki/4Gfy>

Potential queries (ii)

- “Developing practical solutions for monitoring open access and open data policy compliance and tracking impact.”
 - Add queries about
 - Licensing <https://w.wiki/4Gxa>
 - Policies <https://w.wiki/4GxX>
- Topics of first/ latest ORFG-funded papers
- Funders of first/ latest papers on topic X

**FAIR data about
ORFG-funded
research?**



Demographics of ORFG-funded research?



Topics of ORFG-funded research?



Licensing of ORFG-funded research?



**Who cites
ORFG-funded
research?**



**Awards that
correlate with
ORFG-funded
research?**



Reuse of ORFG-funded research?



**Entities that
might be good
fits for ORFG
funding?**



Example profiles:

- NIOSH
- PCORI
- ERC



<https://w.wiki/4GiB>

Stop curating into silos

