



# Persistent Identifiers: Glue and Connections

Patricia Cruse, Executive  
Director, DataCite

Managing Digital Research  
Objects in an Expanding  
Science Ecosystem

# What I am going to talk about



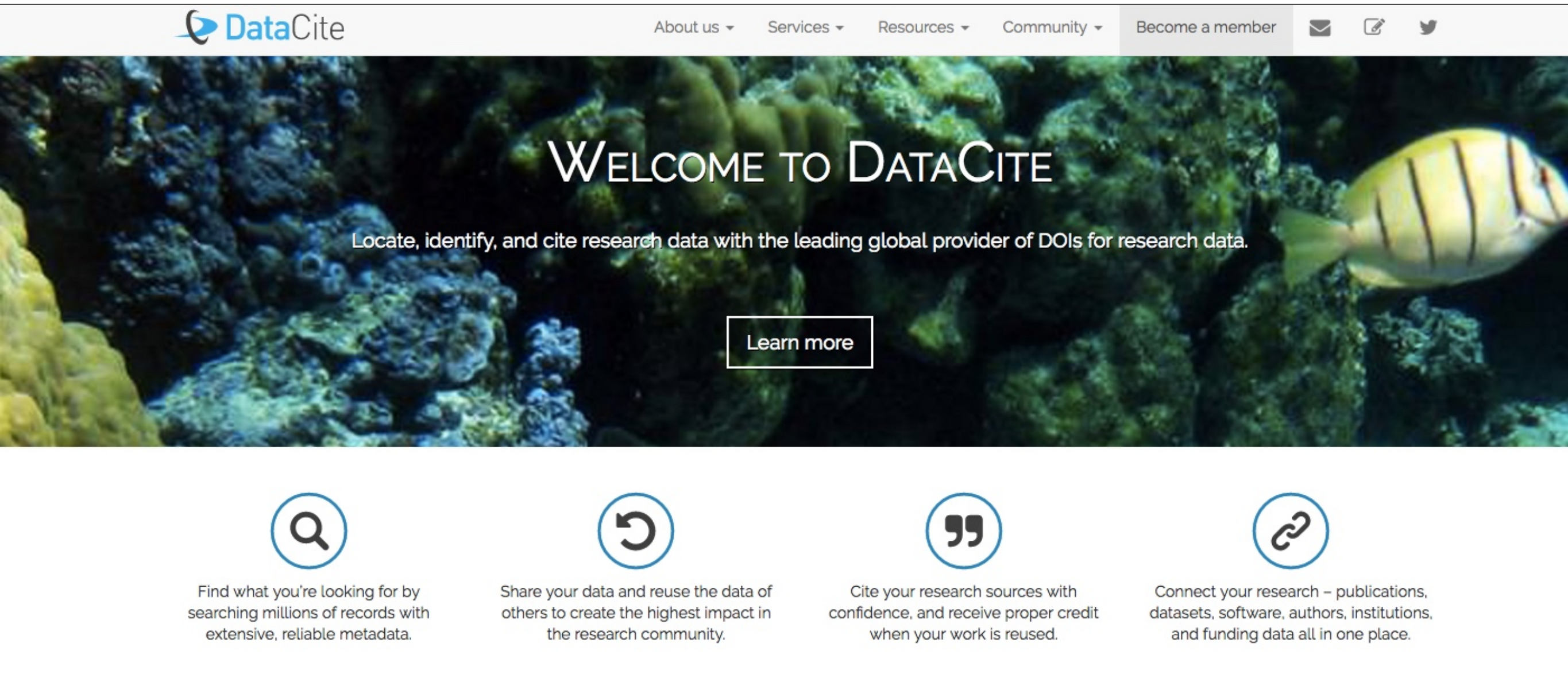
## **Persistent Identifiers = the connective glue**

- What are we trying to solve
- The role of identifiers
- DataCite as a solution
- Making connections
- Usage metrics





Mission: discover, access, use, connect, cite research data



The image shows the top portion of the DataCite website. At the top is a navigation bar with the DataCite logo on the left and links for 'About us', 'Services', 'Resources', 'Community', and 'Become a member' on the right. Below the navigation bar is a large hero section with a background image of a coral reef and a yellow-striped fish. The text 'WELCOME TO DATACITE' is centered in large white letters. Below this, a smaller line of text reads 'Locate, identify, and cite research data with the leading global provider of DOIs for research data.' A 'Learn more' button is centered below the text. At the bottom of the hero section are four circular icons, each with a corresponding description of a DataCite service: a magnifying glass for searching, a circular arrow for sharing and reusing data, quotation marks for citing, and a chain link for connecting research.





**DataCite**

About us ▾ Services ▾ Resources ▾ Community ▾ Become a member

WELCOME TO DATACITE

Locate, identify, and cite research data with the leading global provider of DOIs for research data.

[Learn more](#)

-   
Find what you're looking for by searching millions of records with extensive, reliable metadata.
-   
Share your data and reuse the data of others to create the highest impact in the research community.
-   
Cite your research sources with confidence, and receive proper credit when your work is reused.
-   
Connect your research – publications, datasets, software, authors, institutions, and funding data all in one place.

Not-for-profit global initiative – Member organization  
– Community driven –  
over **1300** data centers – over **9.5** million DOIs







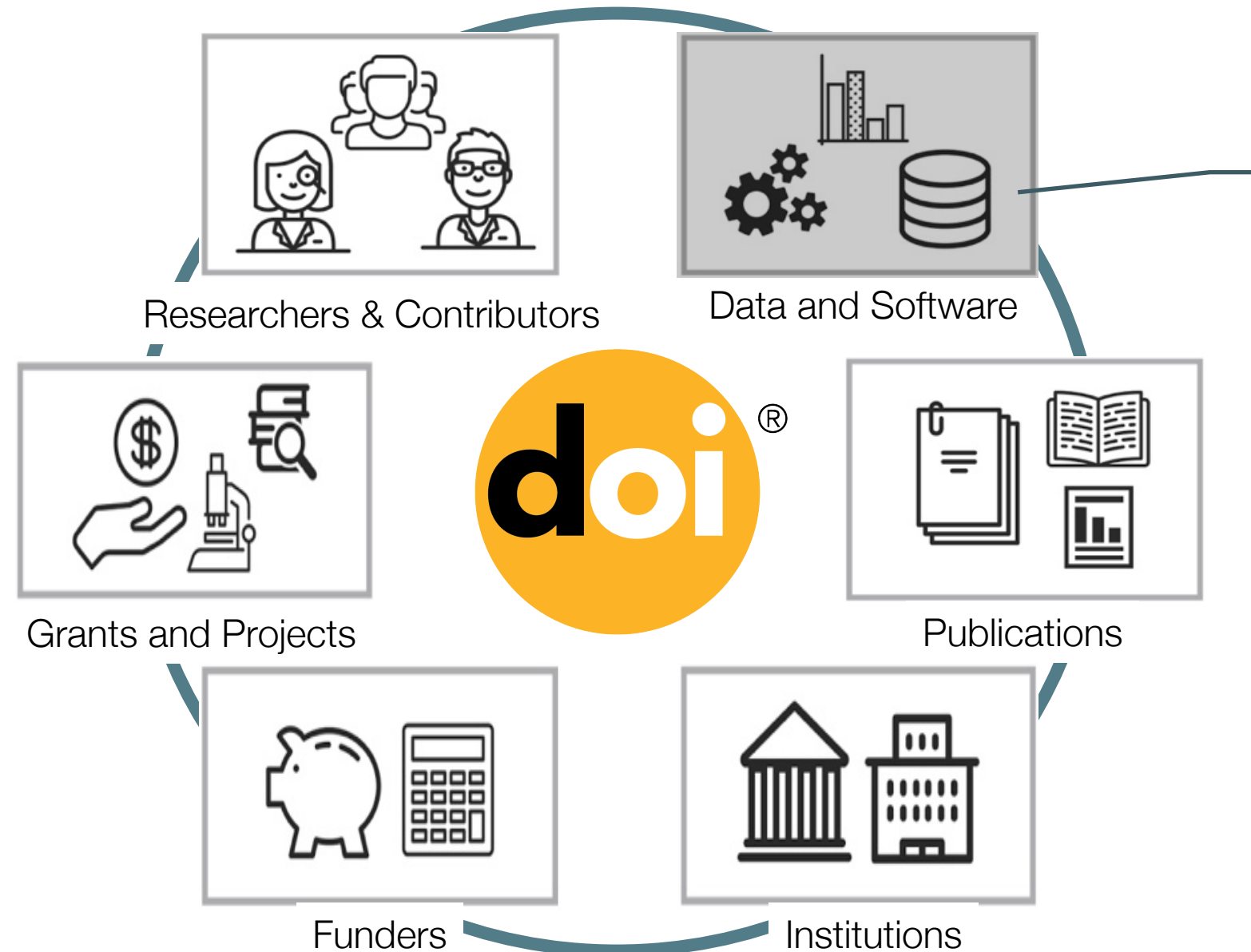


# The community challenges

---

- Get credit and attribution (the village)
- Comply with publishers' data sharing policies 
- Meet funder mandates 
- Meet institutional requirements 
- Respond to community norms and practices 

# Persistent identifiers: connecting research



Data are the primary citable objects

- Connect data to data
  - Versions
  - Physical samples
  - Software
  - Workflows, etc.
- Connect to researchers
- Connect data to publications
- Connect data to funders
- Connect data to *organizations*
- Connect data to *grants and projects*



# DataCite's approach

---



## Provide technical infrastructure:

- Create DOIs for research data
- Build and adopt services that promote data sharing
- Integrate with other community services



## Provide community infrastructure:

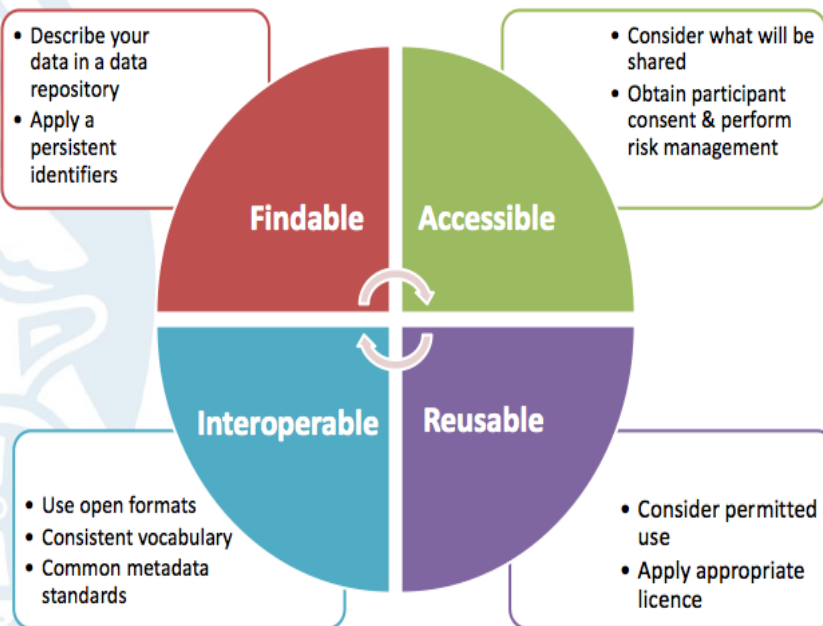
- Advocate & communicate about the importance of data sharing



# Emerging community initiatives



## FAIR data



English

ABOUT COMMUNITY GROUPS RESOURCES NEWS + BLOGS EVENTS PUBLICATIONS MEDIA DONATE

FORCE11 » Joint Declaration of Data Citation Principles - FINAL

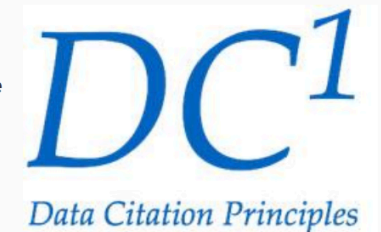
## JOINT DECLARATION OF DATA CITATION PRINCIPLES - FINAL

When citing please use: Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014 <https://doi.org/10.25490/a97f-egyk>

### Endorsement List

#### PREAMBLE

Sound, reproducible scholarship rests upon a foundation of robust, accessible data. For this to be so in practice as well as theory, data must be accorded due importance in the practice of scholarship and in the enduring scholarly record. In other words, data should be considered legitimate, citable products of research. Data citation, like the citation of other evidence and sources, is good research practice and is part of the scholarly ecosystem supporting data reuse.



English

ABOUT COMMUNITY GROUPS RESOURCES NEWS + BLOGS EVENTS PUBLICATIONS MEDIA DONATE

FORCE11 » Software Citation Principles (Published 2016)

## SOFTWARE CITATION PRINCIPLES (PUBLISHED 2016)

This paper has been published in *PeerJ Computer Science*. Please cite it as:  
Smith AM, Katz DS, Niemeyer KE, FORCE11 Software Citation Working Group.  
(2016) Software Citation Principles. *PeerJ Computer Science* 2:e86.  
DOI: [10.7717/peerj-cs.86](https://doi.org/10.7717/peerj-cs.86)

ARFON M. SMITH<sup>1</sup>, DANIEL S. KATZ<sup>2</sup>, KYLE E. NIEMEYER<sup>3</sup>, AND THE FORCE11 SOFTWARE CITATION WORKING GROUP

Corresponding author: Daniel S. Katz<sup>2</sup>, [d.katz@ieee.org](mailto:d.katz@ieee.org).

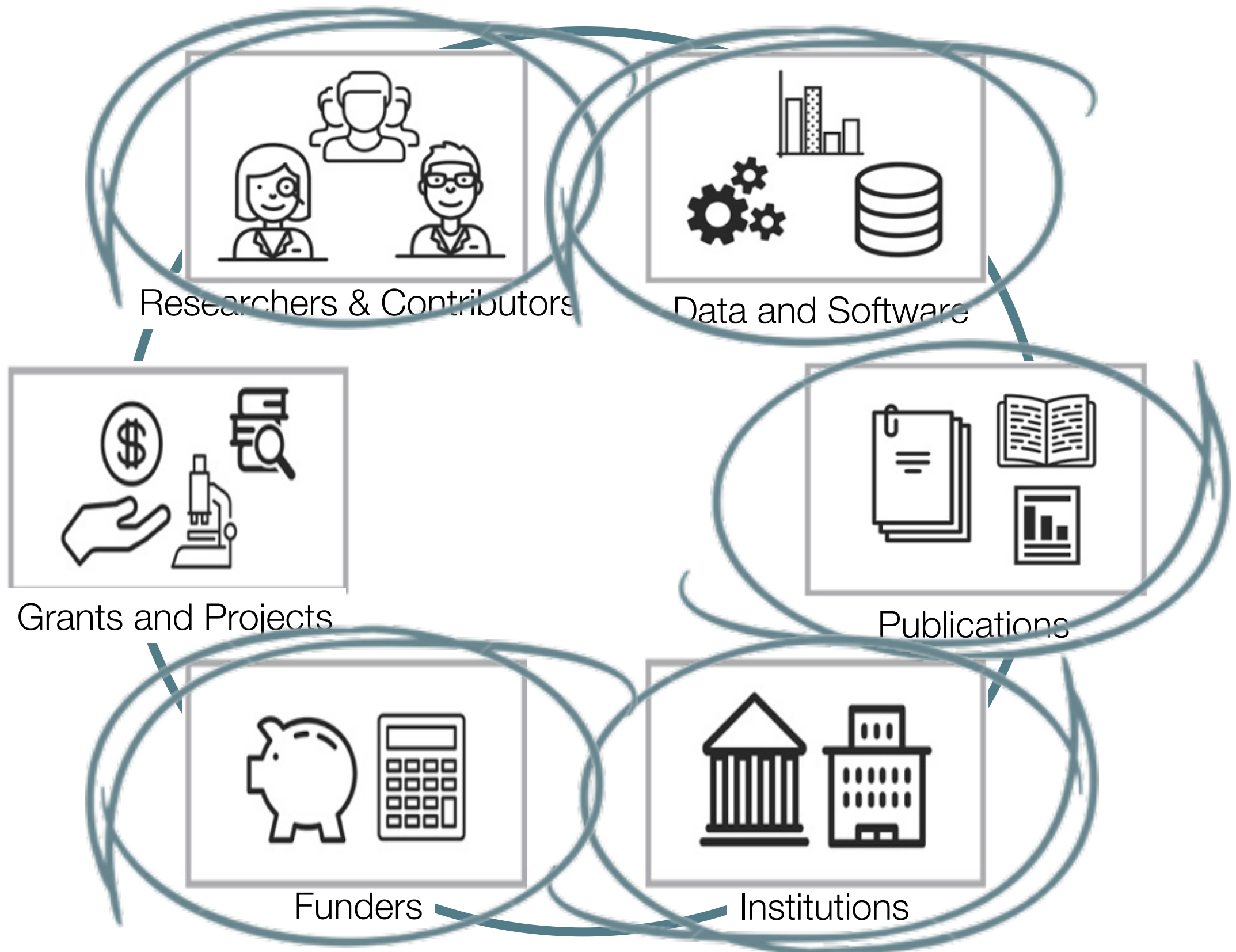
<sup>1</sup>CitHub, Inc., San Francisco, California, United States

<sup>2</sup>National Center for Supercomputing Applications & Electrical and Computer Engineering Department & School of Information Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States

<sup>3</sup>School of Mechanical, Industrial, and Manufacturing Engineering, Oregon State University, Corvallis, Oregon, United States



# Connecting the pieces



# Linking Data with Data



## Why it matters

- provide a complete picture of the data environment
- multiple versions of the same dataset
- subsets of larger datasets
- dynamic data
- software
- workflows



# Linking data to data example



Dataset

DataCite Search

Works

People

Data Centers

## Biom mineralization control related to population density under ocean acidification

Stefano Goffredo, Fiorella Prada, Erik Caroselli, B Capaccioni, Francesco Zaccanti, Luca Pasquini, Paola Fantazzini, Simona Fermani, Michela Reggi, Oren Levy, Katharina Elisabeth Fabricius, Zvy Dubinsky & Giuseppe Falini

Dataset published 2014 via PANGAEA - Data Publisher for Earth & Environmental Science



<https://doi.org/10.1594/PANGAEA.837249>

Cite

Add to ORCID record

### 1 Related Work

## Raw Data - "Biom mineralization control related to population density under ocean acidification"

Stefano Goffredo, Fiorella Prada, Erik Caroselli, Bruno Capaccioni, Francesco Zaccanti, Luca Pasquini, Paola Fantazzini, Simona Fermani, Michela Reggi, Oren Levy, Katharina Fabricius, Zvy Dubinsky & Giuseppe Falini

Fileset published 2014 via Figshare



<https://doi.org/10.6084/M9.FIGSHARE.941061>

Cite

Add to ORCID record

References,  
supplements

# Linking Data with Researchers & Contributors

---



## Why it matters

- Credit and attribution
- Answers who did what
- Link one or more contributors to research output





# Seamless integration with ORCID



If you authorize Crossref and DataCite to update your ORCID record



ORCID



**Researchers:** (1) use ORCID iD when submitting dataset (2) authorize DataCite to update your ORCID record.

**Data centers:** (1) collect ORCID identifiers during submission (2) embed iD in the work and include the iD when submitting to DataCite.

**DataCite:** Upon receipt of data from a data center with a valid identifier, DataCite automatically pushes information to the researcher's ORCID record.



and you add your ORCID to your paper or dataset submission

when your publication gets a DOI, your ORCID record will get updated



**AUTOMATICALLY!**

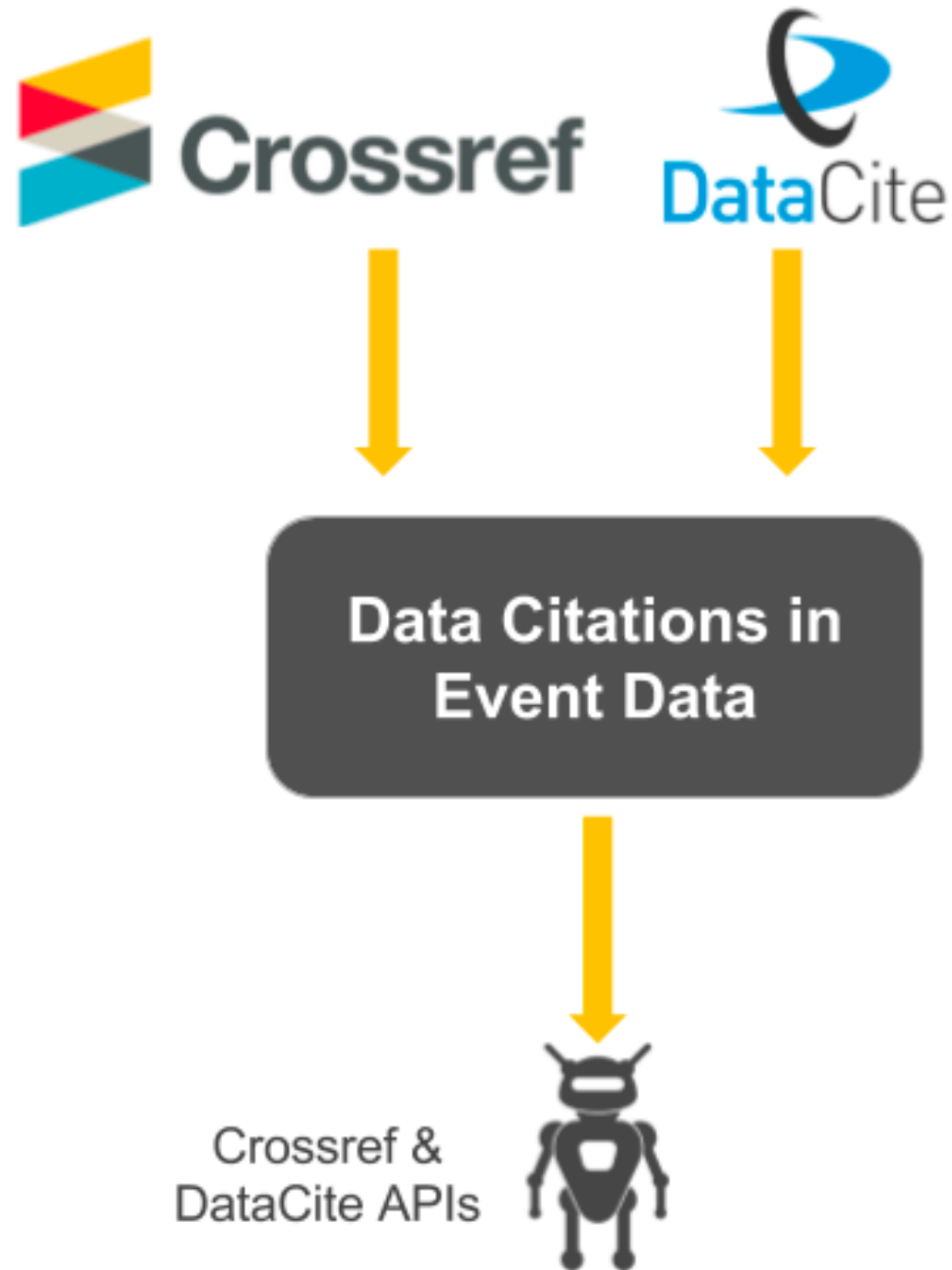
# Linking Data with Articles



## Why it matters

- Increase visibility and discovery of research data and articles
- Place research data in the right context to enable reuse
- Support credit attribution
- Challenges:
  - Data underlying findings are not always fully available
  - Data underlying findings described in a are made available, but hidden in supplementary information
  - Data underlying the findings are available, but not properly linked to/from article

# Event Data = Scholix endpoint





# Example 1: One article links to five datasets



Journal  
article

**DataCite Search** Works Contributors Data Centers Members Sources Sign in

**Temperature-Induced Syntheses, Iodine Elimination, Enantiomers Resolution, and Single-Crystal-to-Single-Crystal Transformation of Imidazole-Co(II) Coordination Polymers with Amino-isophthalic Acid as Co-Ligand**  
Journal article published June 9, 2016

Publisher  
American Chemical Society (ACS)

Share on  
Twitter Facebook

DataCite (Crossref) 5

<http://doi.org/10.1021/ACS.CGD.6B00527> Cite

**Relations 5**

**CCDC 1414538: Experimental Crystal Structure Determination**  
Hui-Fang Zhou, Tian He, Ke-Fen Yue, Yong-Liang Liu, Chun-Sheng Zhou, Ni Yan & Yao-Yu Wang  
Work published 2016 via Cambridge Crystallographic Data Centre

Is supplement to <http://doi.org/10.1021/ACS.CGD.6B00527> DataCite (Crossref)

<http://doi.org/10.5517/CCDC.CSD.CC1JGY86> Cite

**CCDC 1059747: Experimental Crystal Structure Determination**  
Hui-Fang Zhou, Tian He, Ke-Fen Yue, Yong-Liang Liu, Chun-Sheng Zhou, Ni Yan & Yao-Yu Wang  
Work published 2016 via Cambridge Crystallographic Data Centre

Is supplement to <http://doi.org/10.1021/ACS.CGD.6B00527> DataCite (Crossref)

<http://doi.org/10.5517/CCDC.CSD.CC14KRDV> Cite

**CCDC 1484151: Experimental Crystal Structure Determination**  
Hui-Fang Zhou, Tian He, Ke-Fen Yue, Yong-Liang Liu, Chun-Sheng Zhou, Ni Yan & Yao-Yu Wang  
Work published 2016 via Cambridge Crystallographic Data Centre

Is supplement to <http://doi.org/10.1021/ACS.CGD.6B00527> DataCite (Crossref)


**Sources**

☐ DataCite (Crossref) 5

**Relation Types**

☐ Is supplement to 5

Related  
data


**DataCite Search**

[Works](#)
[Contributors](#)
[Data Centers](#)
[Members](#)
[Sources](#)
[Sign in](#)

### Armadillo: a template-based C++ library for linear algebra

Conrad Sanderson & Ryan Curtin

Journal article published June 10, 2016 via JOSS

DataCite (Crossref) <sup>1</sup>

<http://doi.org/10.21105/JOSS.00026> [Cite](#)

**Relations** <sup>1</sup>

### Armadillo C++ Linear Algebra Library

Conrad Sanderson

Work published 2016 via Zenodo

Is cited by <http://doi.org/10.21105/JOSS.00026> [DataCite \(Crossref\)](#)

<http://doi.org/10.5281/ZENODO.55251> [Cite](#)

**Publisher**

The Open Journal

**Share on**

[Twitter](#) [Facebook](#)

**Sources**

☐ DataCite (Crossref) 1

**Relation Types**

☐ Is cited by 1

# Linking data to organizations



## Why it matters

- The research environment is complicated (everyone wants credit)
- Bring together scholarly output with all of the stakeholders
  - Researchers
  - Funders
  - Organizations
  - Grants
  - Projects

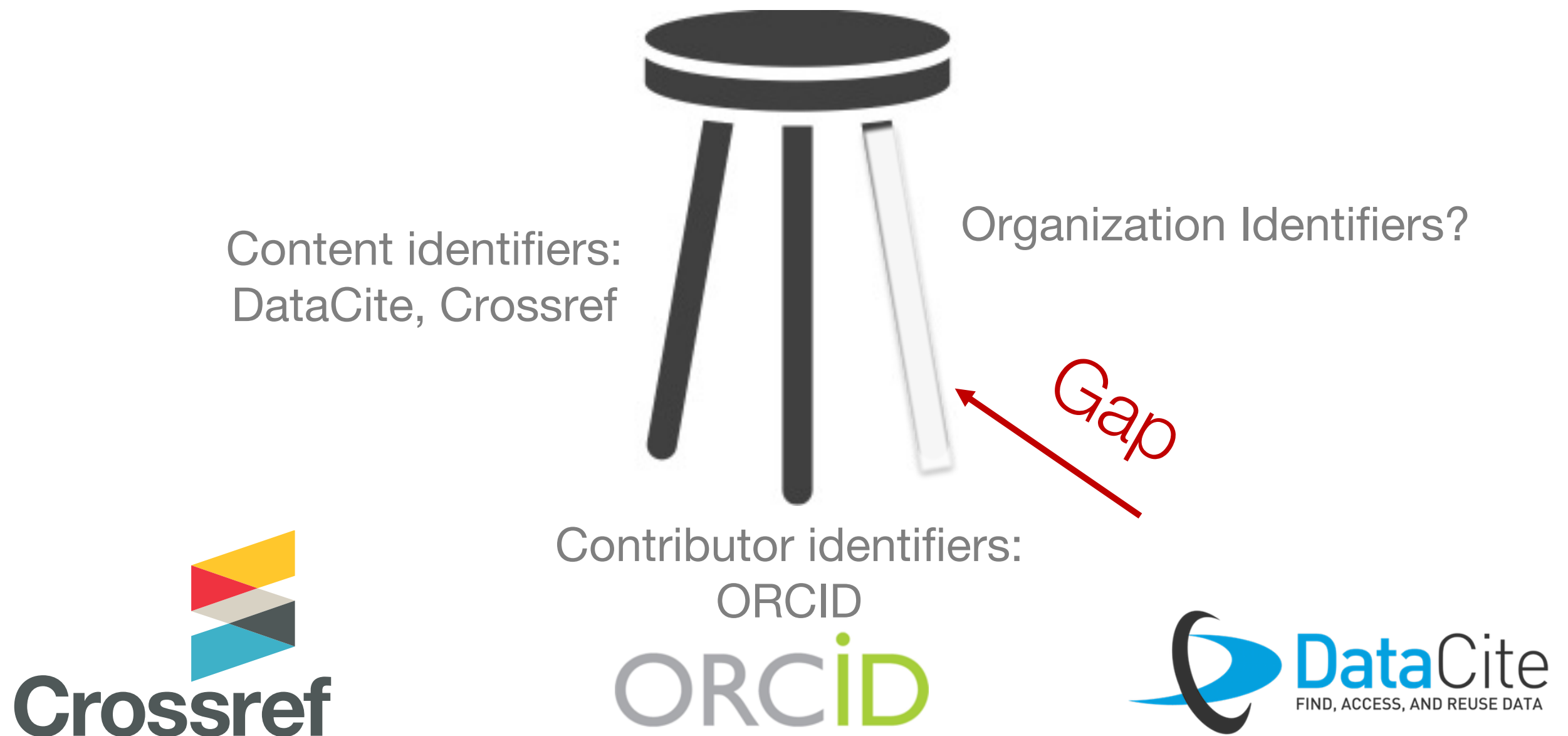




# Linking to organization identifiers



a documented need for a comprehensive, open, and accessible organization identifier infrastructure



# A Community Project



Search



English



Connecting Research  
and Researchers

FOR RESEARCHERS

FOR ORGANIZATIONS

ABOUT

HELP

SIGN IN

## ORGANIZATION ID WORKING GROUP

### WORKING GROUP MEMBERS

#### BREAKOUT GROUP: GOVERNANCE

#### BREAKOUT GROUP: BUSINESS MODEL & FUNDING

#### BREAKOUT GROUP: REGISTRY PRODUCT DEFINITION

## Organization Identifier Working Group

### Summary

The Organization Identifier (OrgID) Working Group was established in January 2017 to refine the structure, principles, and technology specifications for an open, independent, non-profit organization identifier registry to facilitate the disambiguation of researcher affiliations. The scope of work includes three separate but interdependent areas: Governance, Registry Product Definition, and Business Model & Funding. The goal of the Working Group is to create an implementation plan by the end of 2017.

# Connecting with repositories



A searchable catalog of 1,394 research data repositories from around the world in all disciplines, <http://re3data.org>

Many different flavors of data repositories...

- Publisher, e.g., [Dryad](#)
- Sub/Disciplinary, e.g., [RKMP](#)
- Consortium, e.g., [ICPSR](#)
- Country, e.g., [Research Data Australia](#)
- Government, e.g., [Data Portal India](#)
- Research center, e.g., [NASA GES DISC](#)
- Instrument, e.g., [CHANDRA](#)
- General-purpose, e.g., [FigShare](#)
- Roll-your-own, e.g., [DataVerse](#)
- University, e.g., [PURR](#)





NEW



HOME

ABOUT ▾

RESOURCES ▾

ENGAGE

BLOG

GOAL -- develop and deploy the technical and social infrastructure needed to elevate data to a first-class research output alongside more traditional products, such as publications:  
**usage metrics and citations**

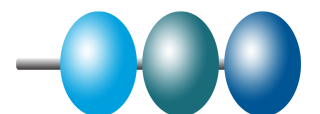
LEARN MORE!

# Make Data Count: a path forward

---



1. Usage metrics and citations
2. Formal recommendation for measuring data usage
3. Develop Hub for all Data Level Metrics (DLM)
4. Make usage tracking easier
5. Drive adoption by showing how it can be done (easily)
6. Engage across all research communities
7. Iterate!



# In sum: connecting research objects

---



## **1. Researchers:**

find, identify, and cite research data and other research objects with confidence

## **3. Journal Publishers:**

enable research articles to be linked to the underlying data/objects

## **2. Data centers:**

provide persistent identifiers for datasets, workflows and best practices for data sharing and citation

## **4. Funders:**

help to track the impact of research funding





**PIDapalooza**

REGISTER NOW!

**Open identifiers deserve their own festival**  
**Join us!**

Girona, Catalonia, Spain  
23rd and 24th January 2018





Thank you!

[patricia.cruse@datacite.org](mailto:patricia.cruse@datacite.org)

<https://www.datacite.org>

Twitter: @datacite