

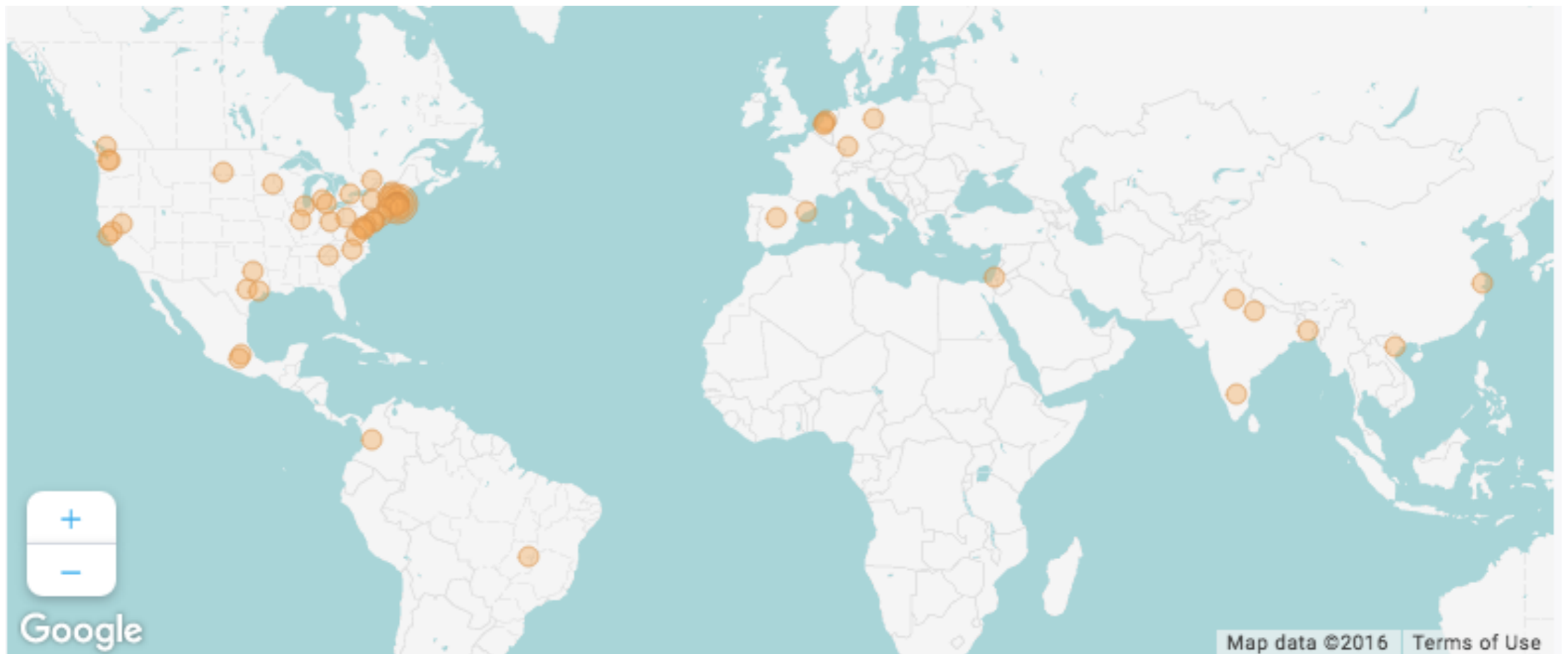
We are the Dataverse  
Community!



#dataverse2016



**185/185** registered



## Day 1 (Monday, July 11)

Dataverse Update Part 1

Dataverse Repositories 1

Using Dataverse Part 1

Using Dataverse Part 2

Reception, Poster Session

## Day 2 (Monday, July 12)

Dataverse Update Part 2

Dataverse Repositories 2

Tools connected with  
Dataverse

Prov

Ops

Training

Closing Remarks, Reception

Prov

## Data Provenance Session

### eeProv: Automatically tracking your data in R

[Margo Seltzer](#), Herchel Smith Professor of Computer Science, Harvard University

Barbara Lerner, Mt. Holyoke College

Emery Boose, Harvard Forest

**Room:** TMEC 426 Conference Room, Tosteson Medical Education Center, Harvard Medical School  
2:30 - 4:00pm, July 12th

Ops

## Dataverse Operational Session

*Kevin Condon, QA Lead, The Dataverse Project and Len Wisniewski, Director of Engineering, IQSS*

**Room:** TMEC 443 Conference Room, Tosteson Medical Education Center, Harvard Medical School  
2:30 - 4:00pm, July 12th

*I've installed Dataverse, now what? Operations policies, practices, and issues.*

Training

## Train the Trainer Session

*Sonia Barbosa, Manager Data Curation, Acquisitions and Archiving, IQSS*

**Room:** TMEC 445 Conference Room, Tosteson Medical Education Center, Harvard Medical School  
2:30 - 4:00pm, July 12th

*"If the only tool you have is a hammer, you tend to see every problem as a nail." – Abraham Maslow*

## Day 3 (Wednesday, July 13)

Developers  
1/2 Day

Privacy Workshop

Interested in  
contributing to the  
Dataverse open-source  
software?

Take the survey:  
<https://goo.gl/ZIT90L>

# Logistics

- Want to take notes during the meeting?  
Public Google Docs, linked from the agenda
- If you have a poster, give it to someone at the registration table or one of the interns
- Transportation to IQSS reception: buses will be ready at 5:15pm



**#dataverse2016**



# The Research Behind Dataverse

Mercè Crosas, Ph.D.  
Chief Data Science and Technology Officer  
Institute for Quantitative Social Science  
Harvard University

 @mercecrosas



Part of international  
initiatives and working  
groups



## RDA/WDS Publishing Data Workflows WG

*Home » Working And Interest Groups » Working Group » RDA/WDS Publishing Data Workflows WG*

WG

### Group details

**Status:** [Recognised & Endorsed](#)

**Chair(s):**

Sunje Dallmeier-Tiessen, Fiona Murphy, Amy Nurnberger, Varsha Khodiyar

**Secretariat Liaison:** Herman Stehouwer

**TAB Liaison:** Rainer Stotzka and Ingrid Dillo

**Case Statement:** [Download](#)

---

Researchers are increasingly encouraged or required to make their research data available for reuse but might often feel there are insufficient incentives for

### RDA/WDS Publishing Data Workflows WG

**Status:** [Recognised & Endorsed](#)

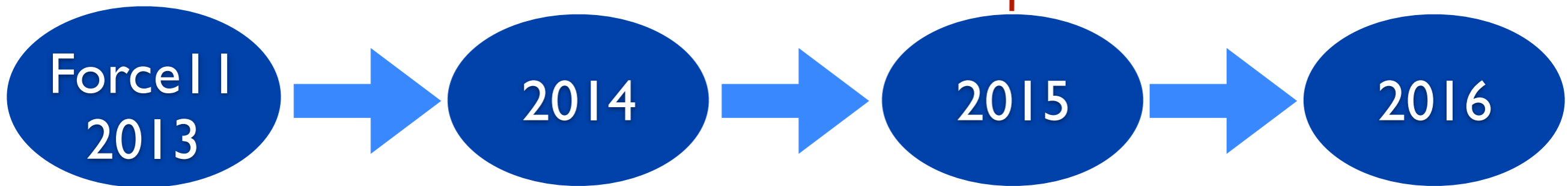
**Chair(s):** Sunje Dallmeier-Tiessen, Fiona Murphy, Amy Nurnberger, Varsha Khodiyar

**Group Email:**

**Secretariat Liaison:** Herman Stehouwer

**TAB Liaison:** Rainer Stotzka and Ingrid Dillo

Starr J, Castro E, Crosas M, Dumontier M, Downs RR, Duerr R, Haak L, Haendel M, Herman I, Hodson S, et al. *Achieving human and machine accessibility of cited data in scholarly publications*. PeerJ Computer Science [Internet]. 2015. Publisher's Version



\$1000 challenge award for proposing the *Amsterdam Manifesto on Data Citation Principles*



Created by a group representing 30 data organizations around the world

Data Citation Implementation Pilot (DCIP)

- Group Home
- Members
- Workshops/Events
- Links & Files
- Google Forum
- Calendar

## EG4 DCIP: REPOSITORY EARLY ADOPTERS

### JUNE 22 WORKSHOP INFO

#### GROUP DESCRIPTION

This group is working with data repositories on best practices that repositories can follow to support data citation:

- machine-readable landing pages
- use of persistent identifiers (in collaboration with the Identifiers Expert Group)
- support of community metadata standards

#### GOALS

Develop a set of recommendations and minimal requirements for repositories to implement a landing page with metadata supporting data citation.

#### PROGRESS/ACTIVITIES

The group has started to have biweekly calls, and will meet in person at the end of June.

#### DELIVERABLES AND TIMELINE

GROUP LEADER

Mercè Crosas



GROUP LEADER

Martin Fenner





**Data Citation Implementation Pilot  
Force I and NIH BioCaddie**

Findable Accessible  
Interoperable and Reusable



## The FAIR Guiding Principles for scientific data management and stewardship

**Mark D. Wilkinson, Michel Dumontier, [IJsbrand Jan Aalbersberg](#), Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons**  Show fewer authors

# FAIR DMP MEMBERS

Displaying 1 - 11 of 11 Active Members

## Working Group on FAIR Data Management Plans



Carole Goble



Daniel S. Katz



John Chodacki



Maryann Martone



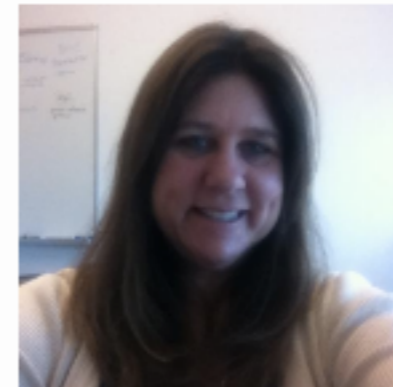
Mercè Crosas



Sarah Jones



Stephanie Simms



Stephanie Hagstrom



Stephen Abrams



Susanna-Assunta



Xiaobin Shen





# Research Grants and Collaborations

# Current Grants



Privacy tools to  
share sensitive data

Data provenance



Alfred P. Sloan  
FOUNDATION

Social Science  
Big Data

Journal articles  
connected to data

Data Privacy

THE LEONA M. AND HARRY B.  
HELMSLEY  
CHARITABLE TRUST

SBGrid and  
Biomedical large-  
scale data



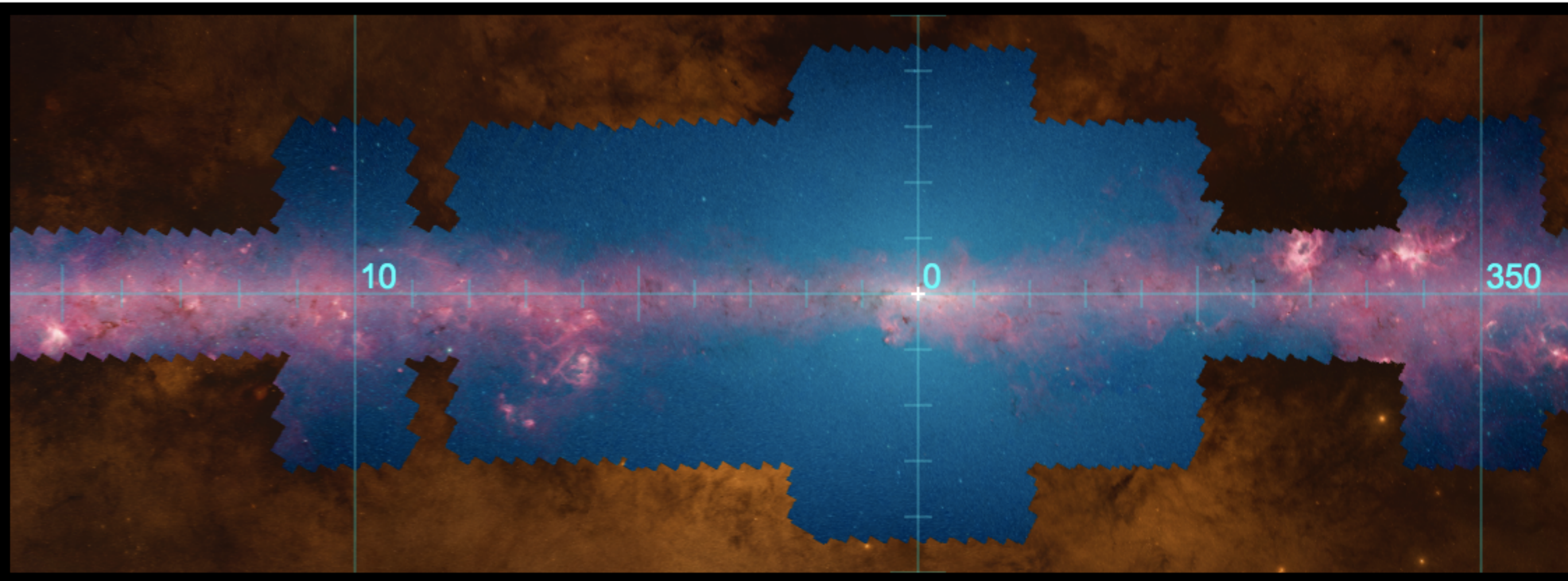
# Integration with the Research Lifecycle





# Milky Way 3D

Galactic Plane Coverage Tool



Extended Observations			Catalogs and Pointed Surveys	
Survey	Wavelength	Co	Spectral Line (2D)	View Region
RAMPS*	1 cm		<input type="checkbox"/>	<input type="checkbox"/>
CORNISH*	60 mm		<input type="checkbox"/>	<input type="checkbox"/>

Each data survey connected to a dataset in Dataverse



Impacting science,  
changing the way we do  
research

# Publishers support sharing data



blog network

ABOUT

Blogs ▾

Topics ▾

Search

[About this](#)

## BioMed Central blog



FOLLOW SUB

## Promoting research data sharing at Springer Nature

This blog was written by Iain Hrynaszkiewicz, Head of Data Publishing at Springer Nature, in support of the newly launched, company-wide Springer Nature Research Data Policies.



Iain Hrynaszkiewicz

Head of Data Publishing at [Springer Nature](#)

# Journal Editors support sharing data

The Profession

---

## How Are We Doing? Data Access and Replication in Political Science

Ellen M. Key, *Appalachian State University*

**ABSTRACT** Data access and research transparency (DA-RT) is a growing concern for the discipline. Technological advances have greatly reduced the cost of sharing data, enabling full replication archives consisting of data and code to be shared on individual websites, as well as journal archives and institutional data repositories. But how do we ensure that scholars take advantage of these resources to share their replication archives? Moreover, are the costs of research transparency borne by individuals or by journals? This article assesses the impact of journal replication policies on data availability and finds that articles published in journals with mandatory provision policies are 24 times more likely to have replication materials than those with no requirements.

**Journals with mandatory data policies, 24 times more likely to have replications materials**



# Research Communities support sharing data

## COPDESS

Coalition on Publishing Data in  
the Earth and Space Sciences

---

[Statement of Commitment](#)

---

[COPDESS Suggested Author  
Instructions and Best Practices  
for Journals](#)

---

[Discussions about COPDESS and  
related efforts](#)

---

[Directory of Repositories](#)

---

[Links to Author & Grantee  
Instructions, Data Policies, and  
Position Statements](#)

---

[COPDESS 2nd Working Meeting,  
20-21 October, Oxford, UK](#)

---

[COPDESS Poster at EGU](#)

Search ...




COPDESS

## Coalition for Publishing Data in the Earth and Space Sciences


*The Coalition for Publishing Data in the Earth and Space Sciences* was formed at a meeting in October 2014 and provides an organizational framework for Earth and space science publishers and data facilities to jointly implement and promote common policies and procedures for the publication and citation of data across Earth Science journals.

# Funding Sources support sharing data

 U.S. Department of Health and Human Services

 National Institutes of Health



Search 

[Data Science Community](#)

[BD2K](#)

[Commons](#)

[News & Events](#)

[About](#)

The Commons supports biomedical discovery by enabling sharing of digital objects

**Some still resist ...**

We are the Dataverse  
Community.

You will eventually  
share your data.

Resistance is futile. 🔥