

DATAVERSE COMMUNITY MEETING

10 Years Sharing Data with Dataverse

[#dataverse2017](https://twitter.com/dataverse2017)

<2006

Once there was the VDC

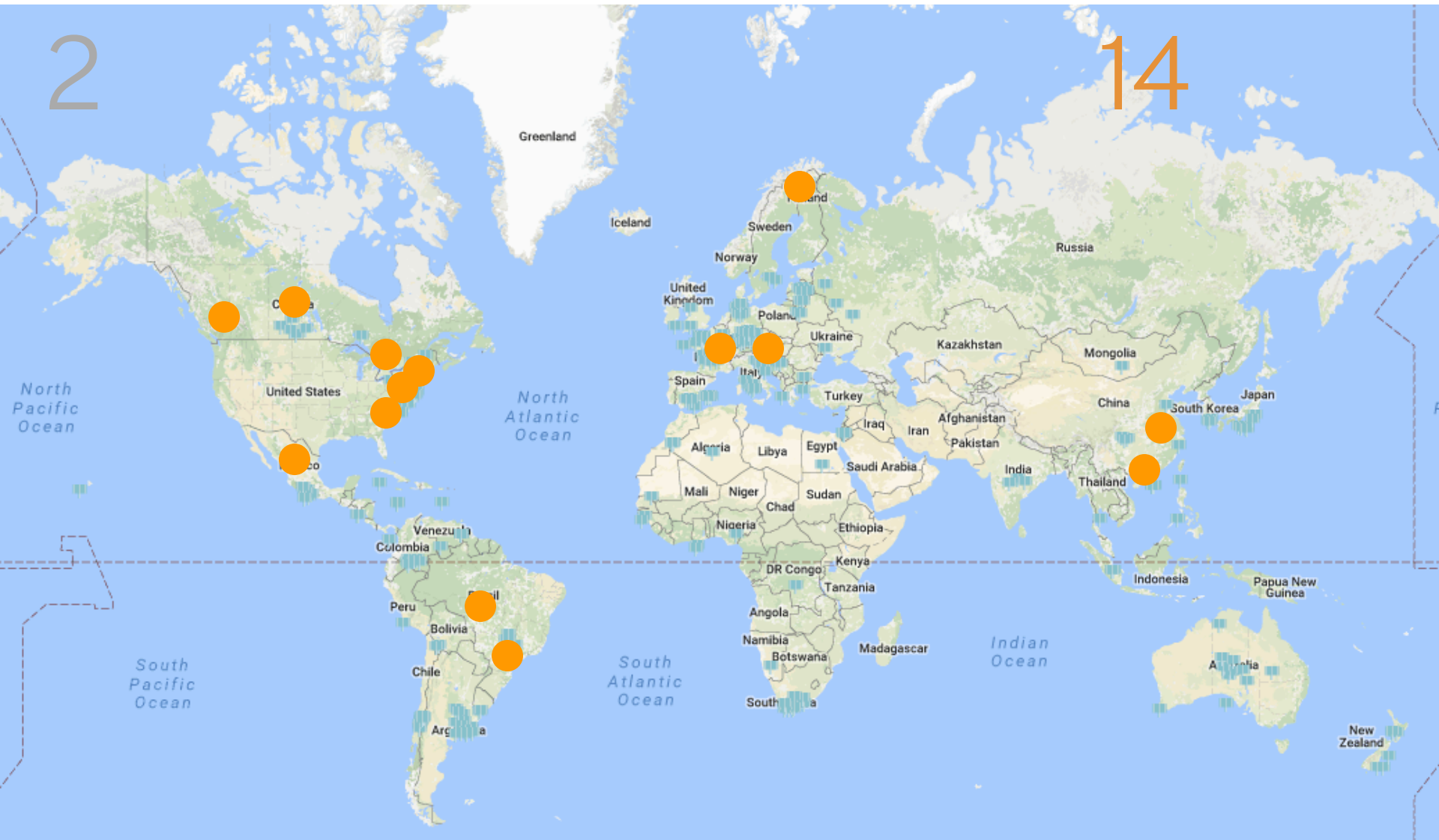
2



2006

2015

And then came the Dataverse Network



2006

2015

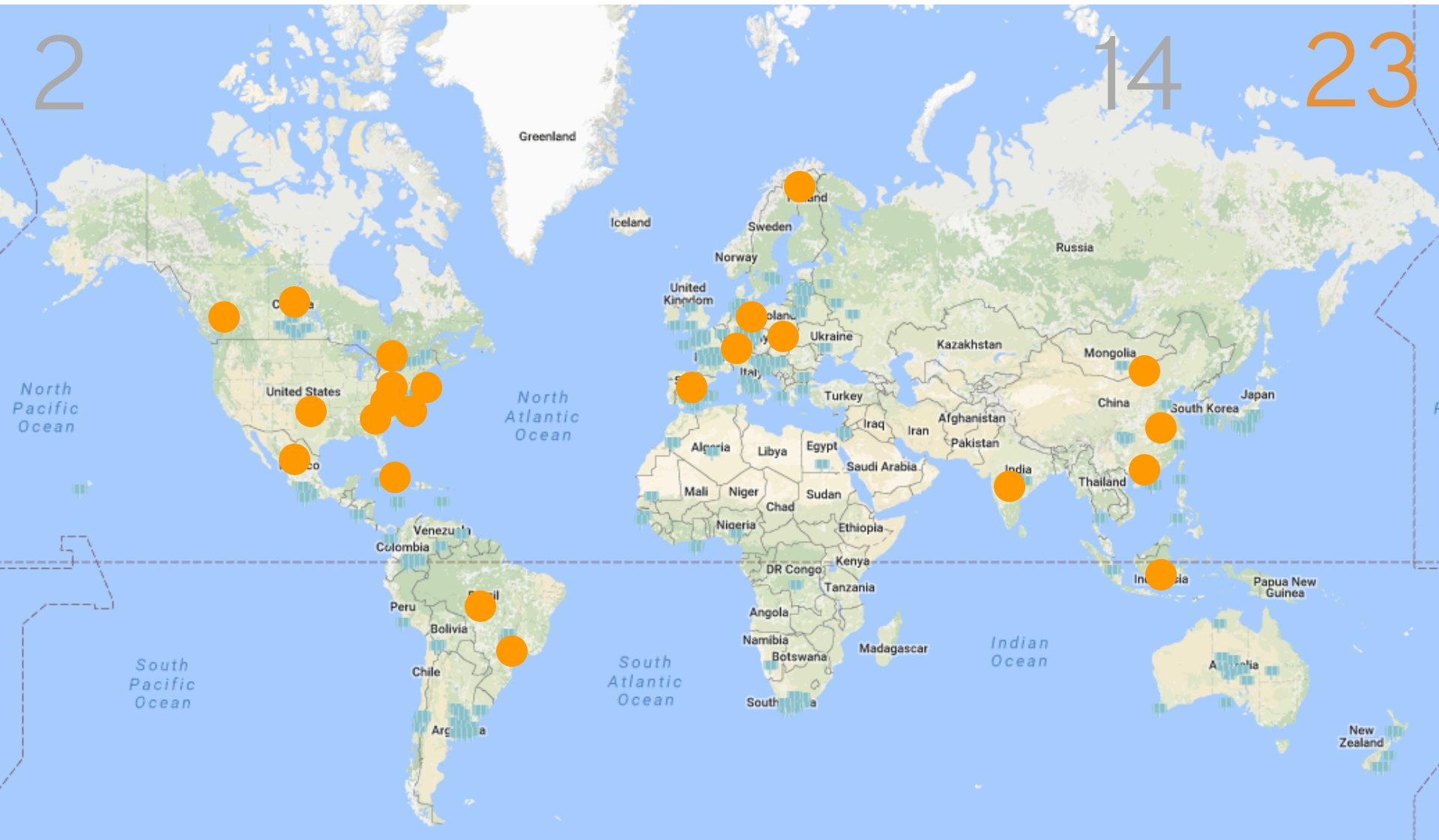
2017

Now we have the Dataverse

2

14

23



RESEARCHERS ARE SHARING AND USING DATA

Harvard Dataverse

> 70,000 datasets
200
datasets/month

> 340,000 files
4,000
files/month

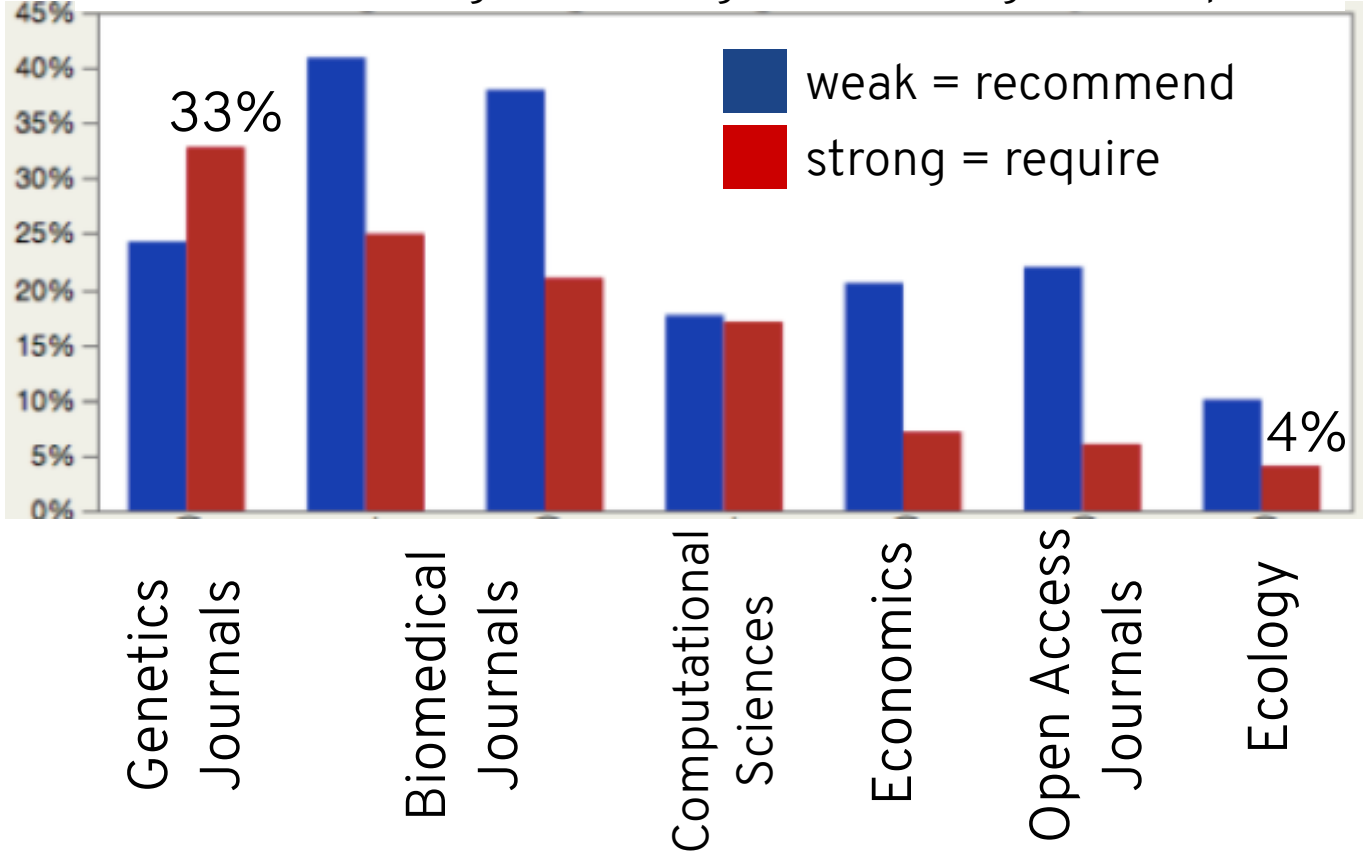
> 2.5 M downloads
60,000
downloads/month

< 2006



When we started, there were **very few** journals with data policies, **no** data requirements from funders

Weak data sharing and strong data sharing vs. disciplines



Now,
Journals
 across
 disciplines
 start
 supporting
 data policies

Castro, Crosas, Garnett, Sheridan, Altman, 2017, Journal of Scholarly Publishing, *Forthcoming*

PRIVATE RESEARCH FUNDERS

- Bill and Melinda Gates Foundation Information Sharing Approach
- Sloan Foundation Data Sharing Policy
- Wellcome Trust Data Sharing Policy
- Arnold Foundation
- Moore Foundation
- Robert Wood Johnson Foundation
- HHMI Policy on the Sharing of Publication-Related Materials, Data and Software

PUBLIC RESEARCH FUNDERS

- Department of Agriculture
- Department of Commerce
- Department of Defense
- Department of Education
- Department of Energy
- Department of Health and Human Services
 - Agency for Healthcare Research and Quality (AHRQ)
 - Assistant Secretary for Preparedness and Response (ASPR)
 - Center for Disease Control and Prevention (CDC)
 - Food and Drug Administration (FDA)
 - National Institutes of Health (NIH)
- Department of Homeland Security
- Department of Housing and Urban Development
- Department of Interior
- Department of Labor
- Department of Transportation
- Department of Veterans Affairs
- Environmental Protection Agency (EPA)

And
Funders
require data
sharing

**WE ARE EXPERIENCING A
CULTURAL CHANGE**

WE ARE EXPERIENCING A
CULTURAL CHANGE

**WE ARE THE CULTURAL
CHANGE!**

**King, 1995, Replication,
Replication**

**2014, Joint Declaration of Data
Citation Principles**

**Wilkinson et al, 2016, The FAIR
Guiding Principles for Scientific
Data Management and
Stewardship**

Altman et al, 2001, A Digital Library for the Dissemination
and Replication of Quantitative Social Science

Pepe et al, 2014, How Do
Astronomers Share Data?

Bierer, Crosas, Pierce, 2017, Data
Authorship as an Incentive to
Data Sharing

Altman and King, 2007, A Proposed for the
Scholarly Citation of Quantitative Data

Goodman et al, 2014, Ten Simple Rules for
the Care and Feeding of Scientific Data

King, 2007, An Introduction to the Dataverse
Network as an Infrastructure for Data Sharing

Crosas, Honaker, King, Sweeney, 2015,
Automating Open Science for Big Data

Crosas, 2012, The Dataverse Network: an open source
application for sharing, discovering, and preserving research
data

Castro et al, 2015, Achieving Human and
Machine Accessibility of Cited Data

Crosas, 2013, A Data Sharing Story

Sweeney, Crosas, Bar-Sinai, 2015, Sharing Sensitive
Data with Confidence: The DataTags System

Altman and Crosas, 2013, The Evolution to Data
Citation: from principles to implementation

Meyer et al. 2016, Data Publication with the
Structural Biology Data Grid Supports Live Analysis

The Dataverse project and team leading many aspects of data sharing

AN ACTIVE TEAM AND COMMUNITY

**METRICS FROM LAST YEAR,
JUNE 2016 TO JUNE 2017**

Community

22 COMMUNITY CALLS

190 ATTENDEES

25 ORGANIZATIONS/UNIVERSITIES

10 COUNTRIES

Community

**975 GOOGLE
GROUP
MESSAGES**

Community

**7,114 IRC
MESSAGES**

245 UNIQUE USERS

IQSS Dataverse Team

12 SPRINTS

(STARTED IN JANUARY 2017)

IQSS Dataverse Team

220 STANDUP MEETINGS

IQSS Dataverse Team

**52,000 SLACK
MESSAGES**

Code

43 GITHUB CONTRIBUTORS

Code

**334 PULL
REQUESTS**

Code

**8,335 GITHUB
COMMITTS**

Support

**1,153 SUPPORT
TICKETS**



DATAVERSE CUP 2017

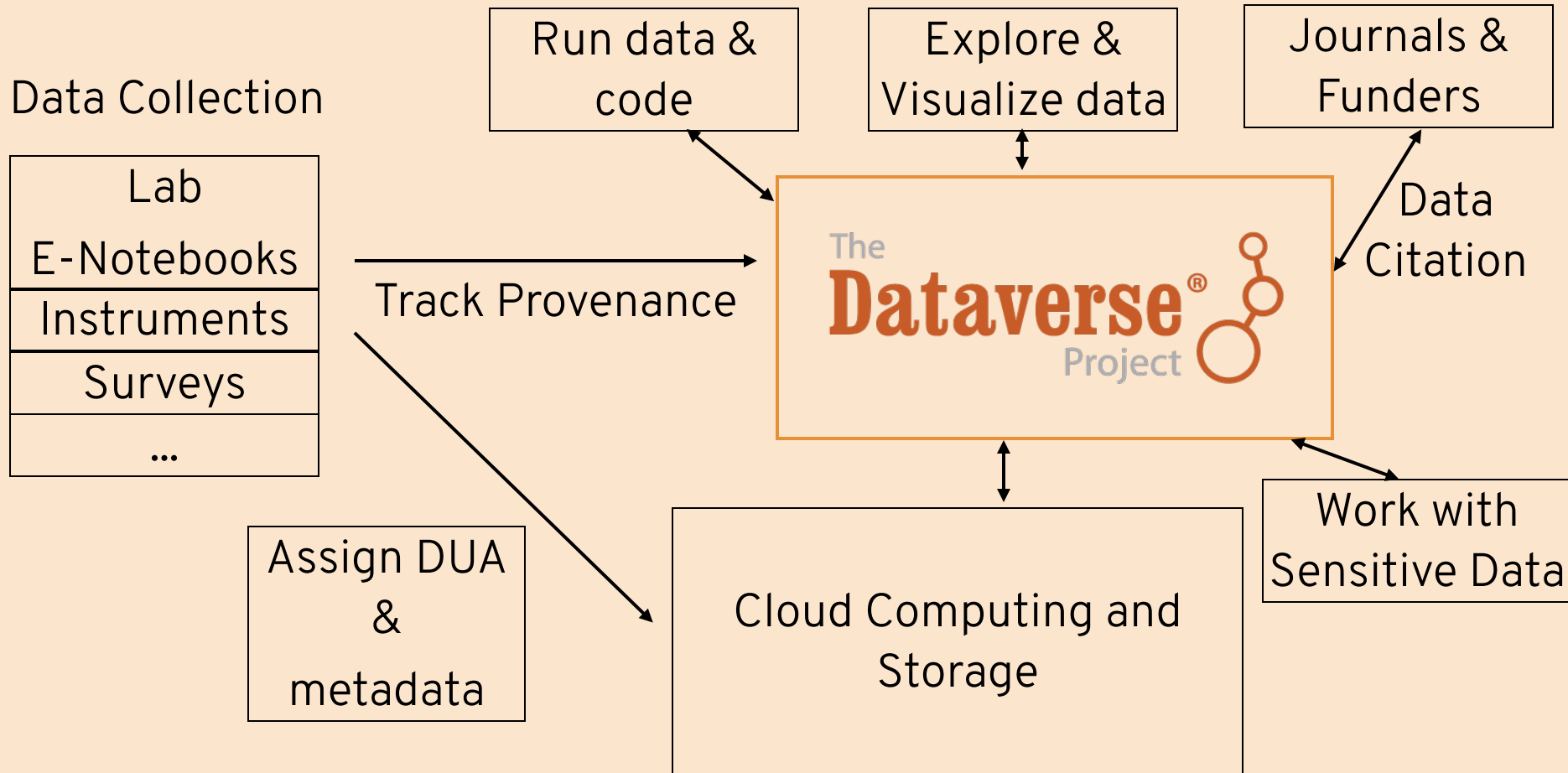


A world map with a light blue background and green landmasses. Numerous small, semi-transparent blue rectangular markers are scattered across the map, representing data points. The markers are most densely clustered in North America, Europe, and East Asia, with fewer markers in South America, Africa, and Australia. The map includes labels for major countries and oceans.

**A VISION:
DATAVERSE AS A KEY PART OF
THE FULL RESEARCH DATA
LIFECYCLE**

TOWARDS A DATA-CENTRIC RESEARCH LIFECYCLE

FROM DATA COLLECTION, TO COMPUTING AND SHARING



COMMUNITY



RESEARCH COLLABORATIONS

Data Privacy
Big Data
Data Policies
Replication
...



INSTITUTIONS REQUIREMENTS

JOURNALS REQUIREMENTS

FUNDERS REQUIREMENTS



STANDARDS AND BEST PRACTICES



TECHNOLOGY ADVANCES

