

UCLA

Presentations

Title

Research Data Alliance in the Science Data-Sharing Landscape

Permalink

<https://escholarship.org/uc/item/8vs1z6s3>

Author

Borgman, Christine L.

Publication Date

2018-12-05

Supplemental Material

<https://escholarship.org/uc/item/8vs1z6s3#supplemental>

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

The Research Data Alliance in the Data Sharing Landscape

Christine L. Borgman

Distinguished Research Professor
Center for Knowledge Infrastructures
University of California, Los Angeles

<http://christineborgman.info>

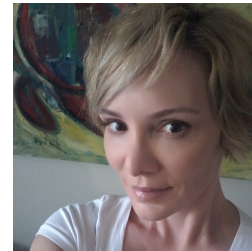
<https://knowledgeinfrastructures.gseis.ucla.edu>

@scitechprof

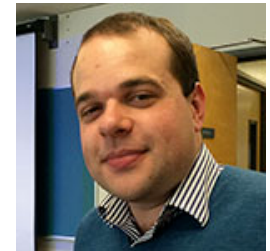
Research Data Alliance, France
National Open Science Plan for France:
From Strategy to Action
Paris, 5 December 2018



Christine Borgman



Bernie Boscoe



Peter Darch



Milena Golshan



Irene Pasquetto



Michael Scroggins



Cheryl Thompson



Morgan Wofford



NATIONAL PLAN FOR OPEN SCIENCE

4TH JULY 2018

#openscience

 esr.gouv.fr



- Generalize open access to publications
- Structure research data and make it available through open access
- Be part of a sustainable European and international open science dynamic



MEMBERSHIP

Members: **7580**

Becoming a member of RDA is simple and open to both individuals and organizations

[Register now](#)

WORKING GROUPS

Groups: **95**

Discover what RDA Working and Interest Groups and all other Groups are up to and find out how to join them. [Explore Groups](#)



Building the social and technical bridges to enable open data sharing



GUIDING PRINCIPLES

◎ OPENNESS

Membership is open to all interested individuals who subscribe to the RDAs Guiding Principles. RDA community meetings and processes are open, and the deliverables of RDA working Groups will be publicly disseminated.

◎ CONSENSUS

The RDA moves forward by achieving consensus among its membership. RDA processes and procedures include appropriate mechanisms to resolve conflicts.

◎ BALANCE

The RDA seeks to promote balanced representation of its membership and stakeholder communities.

◎ HARMONIZATION

The RDA works to achieve harmonization across data standards, policies, technologies, infrastructure and communities.

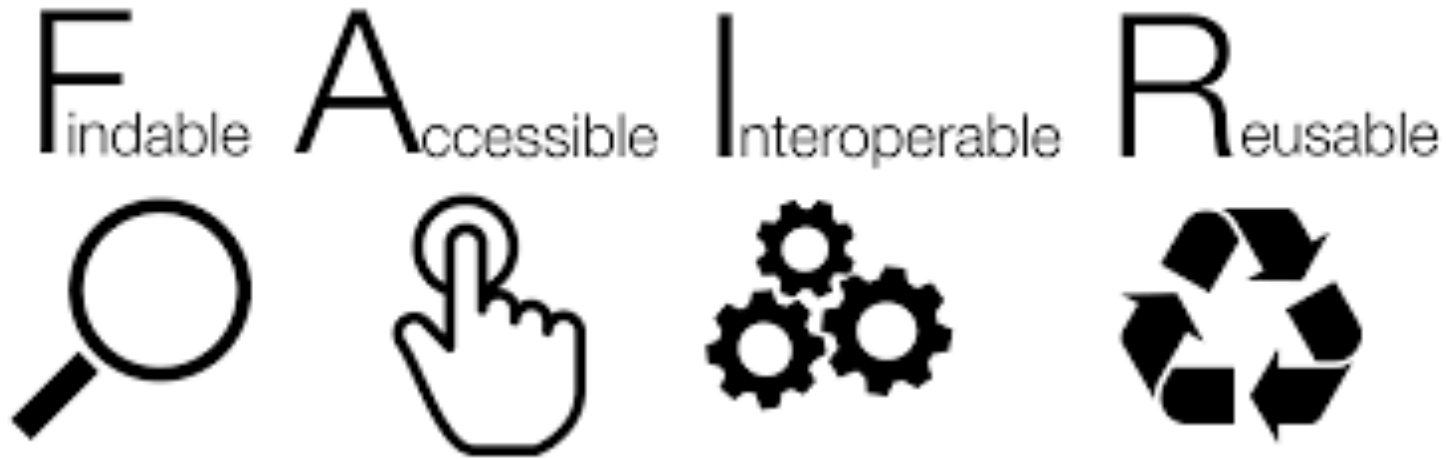
◎ COMMUNITY - DRIVEN

The RDA is a public, community-driven body constituted of volunteer members and organizations, supported by the RDA Secretariat.

◎ NON-PROFIT

RDA does not promote, endorse, or sell commercial products, technologies or services.

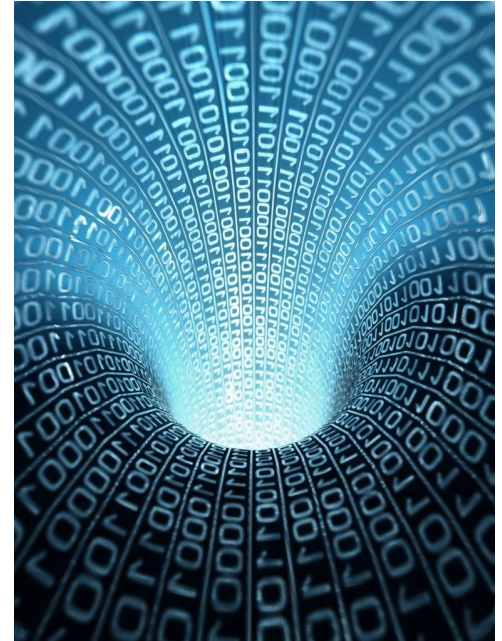
Principles for Data Management



Wilkinson, et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, <http://dx.doi.org/10.1038/sdata.2016.18>

Principles for Access to Research Data

Openness, flexibility, transparency, legal conformity, protection of intellectual property, formal responsibility, professionalism, interoperability, quality, security, efficiency, accountability, and sustainability.



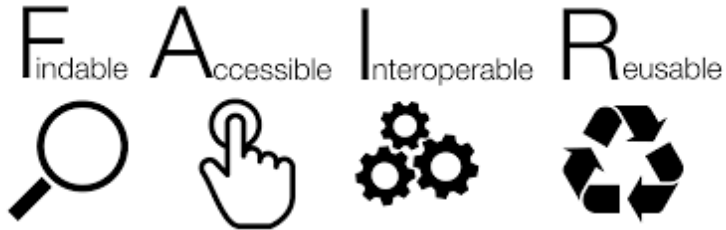
Organization for Economic Cooperation and Development. (2007).
OECD Principles and Guidelines for Access to Research Data from Public Funding.
<http://www.oecd.org/dataoecd/9/61/38500813.pdf>

open by design

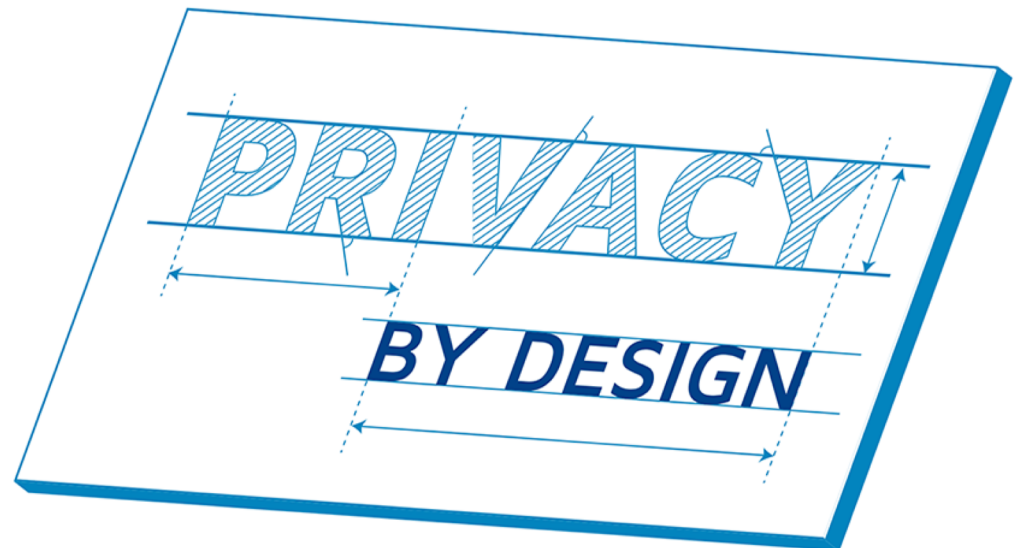
<http://democracyos.eu/blog/open-by-design>



<https://wwwdb.inf.tu-dresden.de/opendatasurvey/>



Wilkinson, et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, <http://dx.doi.org/10.1038/sdata.2016.18>



<https://privacybydesign.foundation/en/>

Borgman, C. L. (2018). Open Data, Grey Data, and Stewardship: Universities at the Privacy Frontier. *Berkeley Technology Law Journal*, 33(2), 365–412. <https://doi.org/10.15779/Z38B56D489>

Data

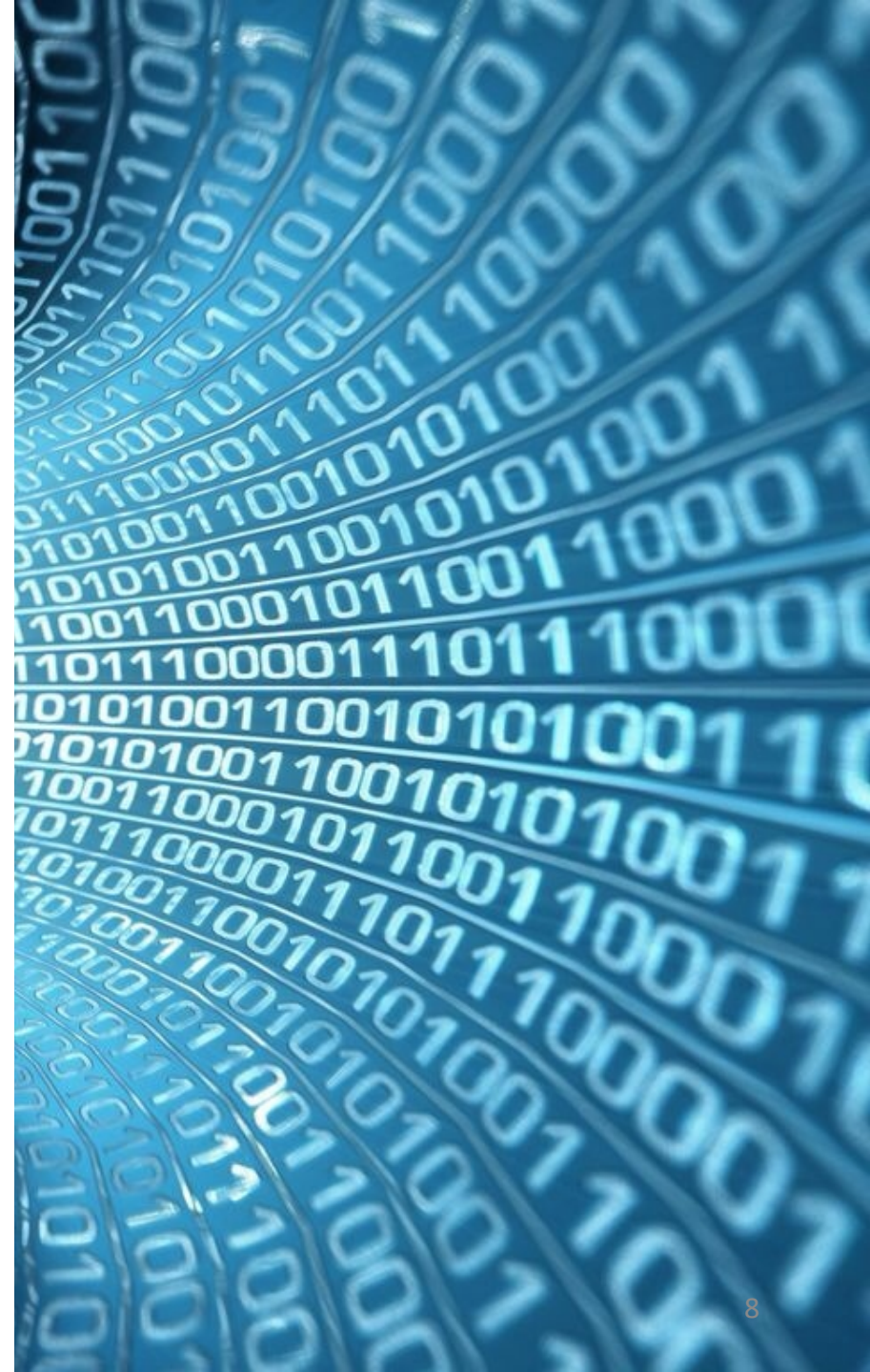
If you can't protect it,
don't collect it.

(privacy and security aphorism)

Therefore:

If you collect it, you
must protect it.

Borgman, C. L. (2018). Open Data, Grey Data, and Stewardship: Universities at the Privacy Frontier. *Berkeley Technology Law Journal*, 33(2), 365–412. <https://doi.org/10.15779/z38B56D489>



Lack of incentives to share data

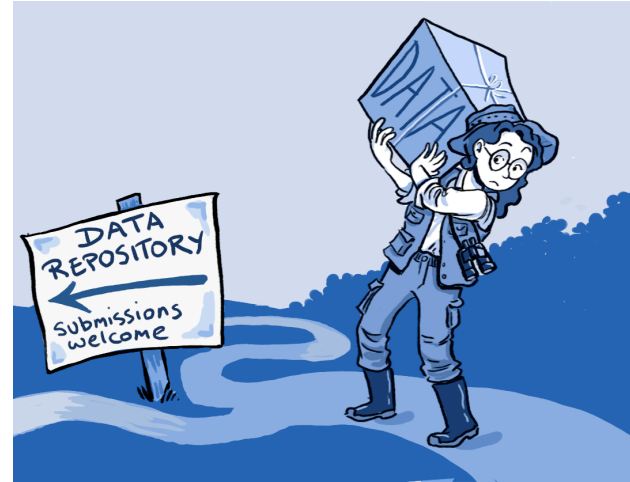


Image source: www.buildingsrus.co.uk/.../target1.htm

- Labor to document data
- Benefits to unknown others
- Competition
- Control
- Confidentiality
- Lack of expertise and staff
- Lack of sustainability...

Data Sharing: End or Means?

- Manage data
- Reuse data
 - Comparison, calibration
 - New inquiries
 - Reproduce research
- Promote transparency in research
- Preserve evidence, models, records
- Promote data standards and interoperability
- “Structure research data and make it available through open access” ...



Organizational choices

- Convening body or standards body?
- Individual or institutional membership?
- Professional society or governing mechanism?
- Independent organization or partnering entity?
- Sustainability mechanisms?





ACTIVE GROUPS

[Data Exchange Working Group](#)[Permissions & Obligations Expression WG](#)[Spatial Data on the Web Working Group](#)[RDF Data Shapes Working Group](#)[Data on the Web Best Practices Working Group](#)[Semantic Web Interest Group](#)[Semantic Web Health Care and Life Sciences Interest Group](#)

NEARBY

[Data Activity Blog](#)[Web of Things - Linked Data as the basis for countering fragmentation of the IoT](#)[The Digital Publishing Activity](#)[The XML Activity](#)[W3C study of practices and tooling for Web data standardisation](#)[W3C Workshop on Privacy and Linked Data](#)

W3C DATA ACTIVITY *Building the Web of Data*

More and more Web applications provide a means of accessing data. From simple visualizations to sophisticated interactive tools, there is a growing reliance on the availability of data which can be “big” or “small”, of diverse origin, and in different formats; it is usually published without prior coordination with other publishers — let alone with precise modeling or common vocabularies. The Data Activity recognizes and works to overcome this diversity to facilitate potentially Web-scale data integration and processing. It does this by providing standard data exchange formats, models, tools, and guidance.

The overall vision of the Data Activity is that people and organizations should be able to share data as far as possible using their existing tools and working practices but in a way that enables others to derive and add value, and to utilize it in ways that suit them. Achieving that requires a focus not just on the interoperability of data but of communities.

W3C gratefully acknowledges support from the European Commission for participation in a number of projects, e.g. [Create-IoT](#), [Big Data Europe](#) (Linked Data), [Boost 4.0](#) (Big data in Industry 4.0) and [SPECIAL](#) (Linked Data for data privacy management).

News

W3C is pleased to announce the First Public Working Draft for the [Data Catalog Vocabulary \(DCAT\) – revised edition](#). DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web. This revised version of DCAT was developed by the [Dataset Exchange Working Group](#) in response to a new set of Use Cases and Requirements based upon extensive experience with the original DCAT specification and related work on DCAT application profiles.

Dave Raggett gave a [plenary presentation on the Web of Things](#) at the opening session of the [FIWARE Summit](#), and later met with [FIWARE Foundation](#) staff to discuss potential opportunities for collaboration between W3C and FIWARE in respect to alignment between the W3C Web of Things object model and API with the FIWARE Orion context broker, which is based upon ETSI's NGSI-LD as a REST API using JSON-LD for querying, updating and notifications of changes to the context, including IoT devices. FIWARE is a leading open source IoT platform.

W3C held a [Workshop on Privacy and Linked Data](#) in Vienna on 17-18 April 2018. The presentations and meeting minutes will be available from the [Workshop page](#).

As a starting point for making W3C a more effective, more welcoming and sustainable venue for communities seeking to develop Web data standards and exploit them to create value added services, we are pleased to announce a [W3C study on Web data standardization](#) that has been produced with support from the [Open Data Institute](#) and [Innovate UK](#).

W3C took part in the January 2018 kick off meeting for the [Boost 4.0](#) European project on big data in smart manufacturing (Industry 4.0). [Our](#) role focuses on standardisation, data governance and certification.

Software is fragile

unlike words carved in stone it can
be deleted or get corrupted



We are building the universal software archive



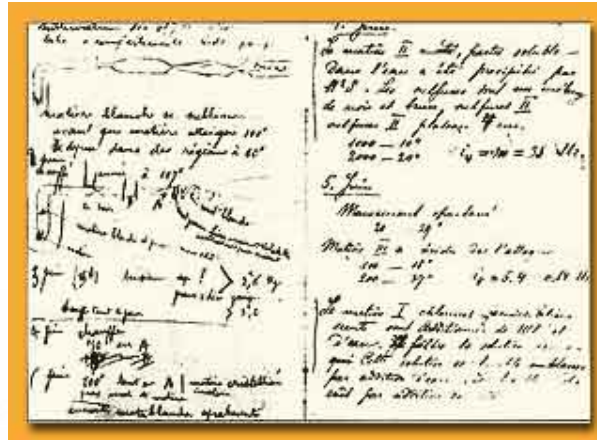
We **collect** and **preserve** software in source code form. because software



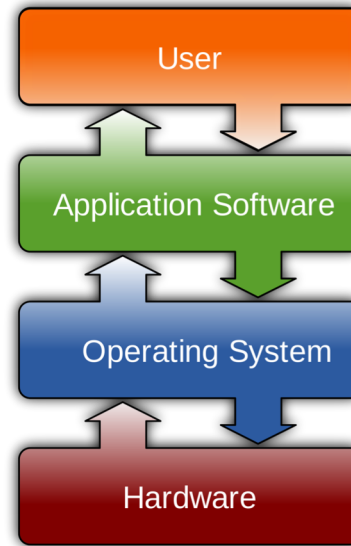
Research Data



Alibaba.com



Marie Curie's notebook aip.org



Wikipedia.org



Pisa Griffin

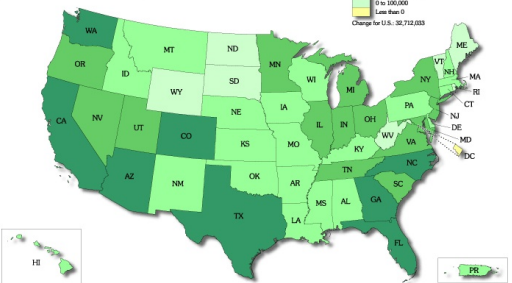


hudsonalpha.org



Figure 2. Numeric Change in Resident Population for the 50 States, the District of Columbia, and Puerto Rico: 1990 to 2000

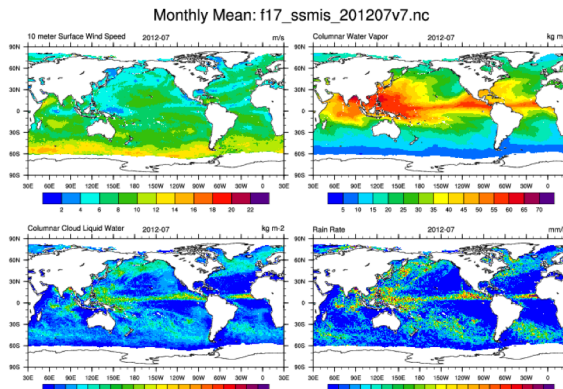
Change in Number of People
 Greater than 1 million
 500,000 to 1 million
 100,000 to 500,000
 0 to 100,000
 Less than 0
 Change for U.S.: 32,712,283



Powered by Geography Division

USCENSUS BUREAU
 Supply the Data. Advance the Nation.

<http://www.census.gov/population/cen2000/map02.gif>



ncl.ucar.edu

Date: 1/2.07.75 Place: Sakaltutan
Zafor

He will grow old in his present house; new house is for sons - 5 sons. Not sure they want to live in village. He will only build another if they want him to. eS came from Germany and did the plastering. He arranged the carpentry in Kayseri. Çok para gitti. (much money went) Has a tractor.

Date: July 1980 Place: Sakaltutan
Zafor:

Household now Zafor and wife; Nazif Unal and wife and youngest son, still a boy. They run two dolmuş; one with a driver from Süleymanlı. Goes in and out once a day. He gets 8,000 a month. Zafor then said, keskin deOil. (not sharp - i.e.: not profitable) I said he did very well on 8,000 TL with only two journeys a day. Nazif Unal has "bought" a Durak (dolmuş stop) from Belediye and works all day in Kayseri.

http://onlineqda.hud.ac.uk/Intro_QDA/Examples_of_Qualitative_Data.php

Data reuse in practice

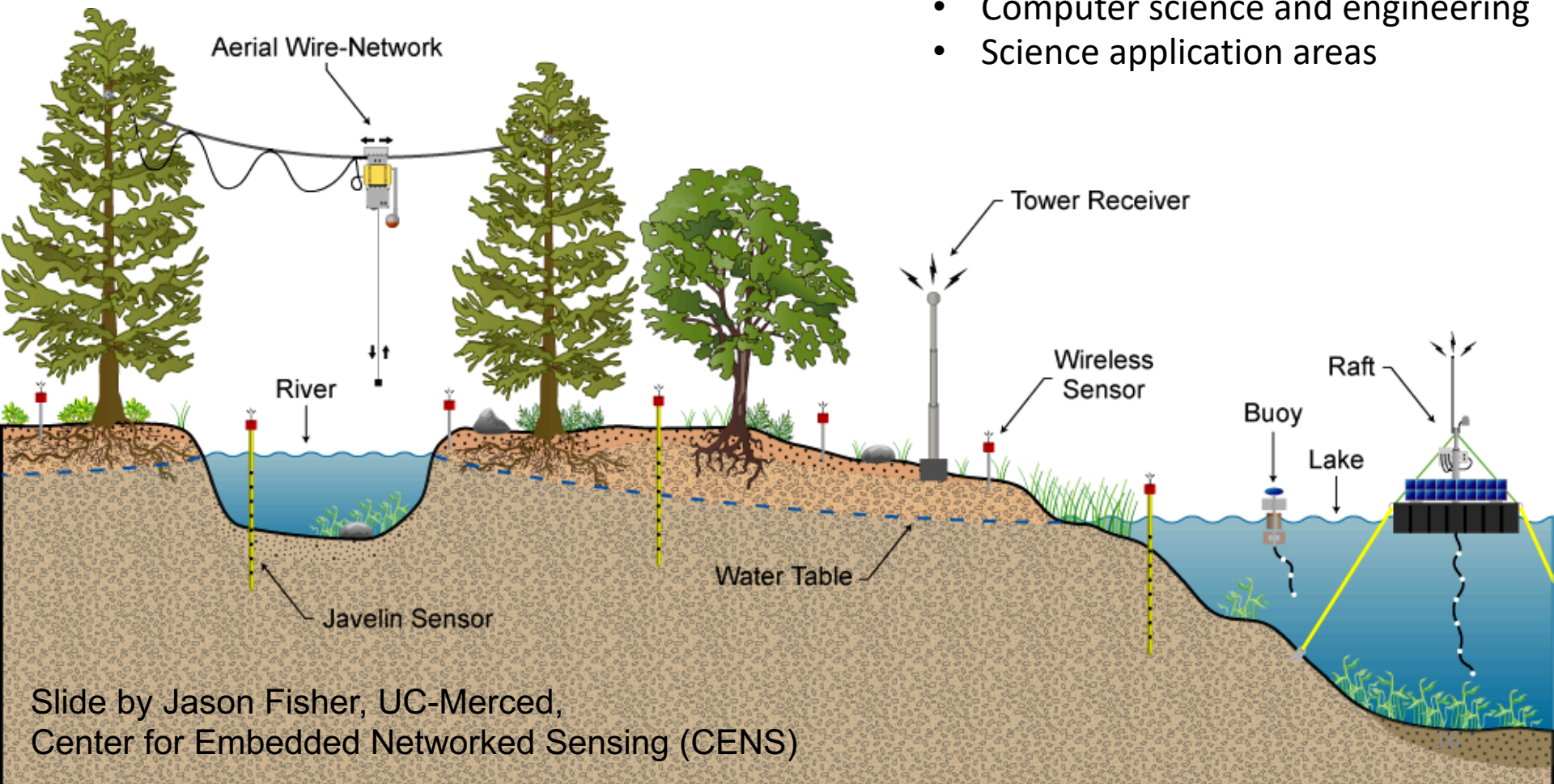
Current research

Center for Knowledge Infrastructures

UCLA

Center for Embedded Networked Sensing

- NSF Science & Tech Ctr, 2002-2012
- 5 universities, plus partners
- 300 members
- Computer science and engineering
- Science application areas



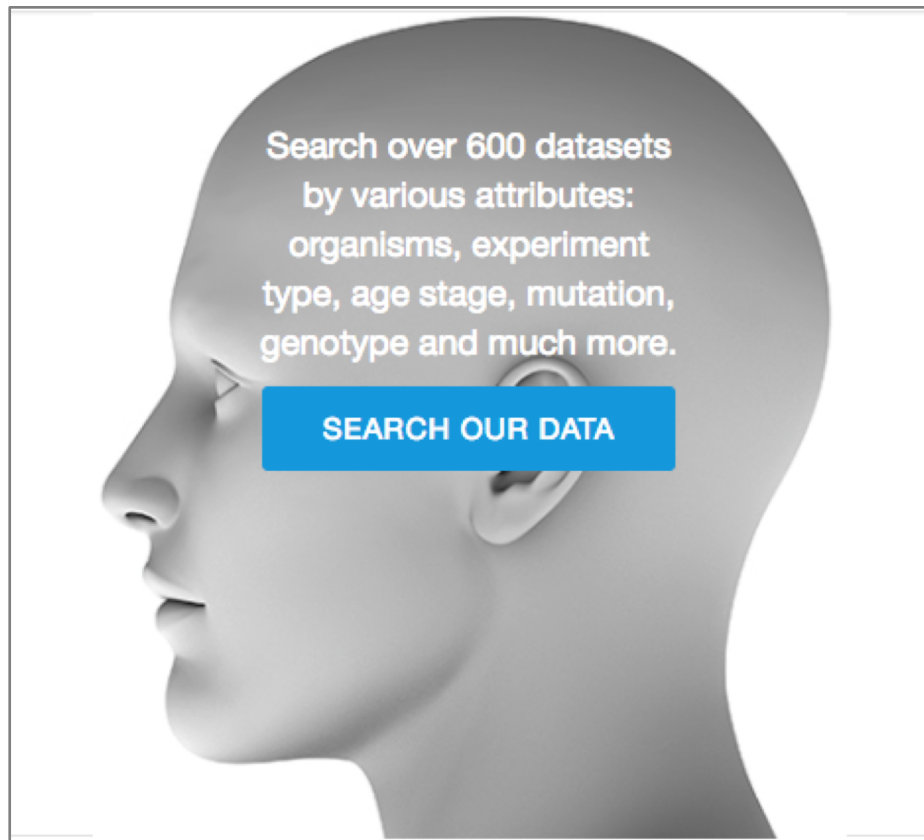
Slide by Jason Fisher, UC-Merced,
Center for Embedded Networked Sensing (CENS)

THE DATAFACE CONSORTIUM

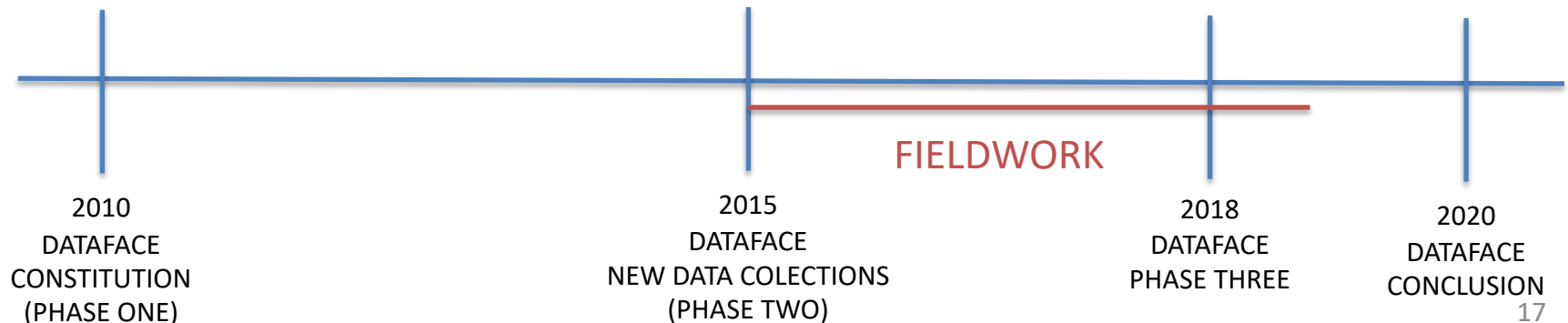
GOAL:

Create a “data resource” that can be mined indefinitely to extract novel knowledge.

- Over 600 “hypothesis free” genomics craniofacial datasets (DNA and facial images)
- All datasets deposited in a open repository **prior to publication.**



Pasquetto, I.V. (2018). *From Open Data to Knowledge Production: Biomedical Data Sharing and Unpredictable Data Reuses*. PhD Thesis. <https://escholarship.org/uc/item/1sx7v77r>



Background and foreground reuses of data

	Background Reuse	Foreground Reuse
Goal of data reuse	“Ground truthing:” calibrate, compare, confirm	Analysis: identify patterns, correlations, causal relationships
Example of data reuse	Instrument calibration, sequence annotation, review summary-level data	Meta-analyses, novel statistical analyses
Frequency of data reuse	Frequent, routine practice	Rare, emergent practice

- Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2018, in review). The Who, What, When, and Why of Reusing Data in Scientific Practice. *Harvard Data Science Review*.
- Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLOS ONE*, 8(7), e67332. <https://doi.org/10.1371/journal.pone.0067332>
- Pasquetto, I.V. (2018). *From Open Data to Knowledge Production: Biomedical Data Sharing and Unpredictable Data Reuses*. PhD Thesis. <https://escholarship.org/uc/item/1sx7v77r>

Background and foreground reuses of data

	Background Reuse	Foreground Reuse
Goal of data reuse	“Ground truthing:” calibrate, compare, confirm	Analysis: identify patterns, correlations, causal relationships
Example of data reuse	Instrument sequence review data level	Meta-analysis statistical
Frequency of data reuse	Frequent routine practice	Rare, emergent practice

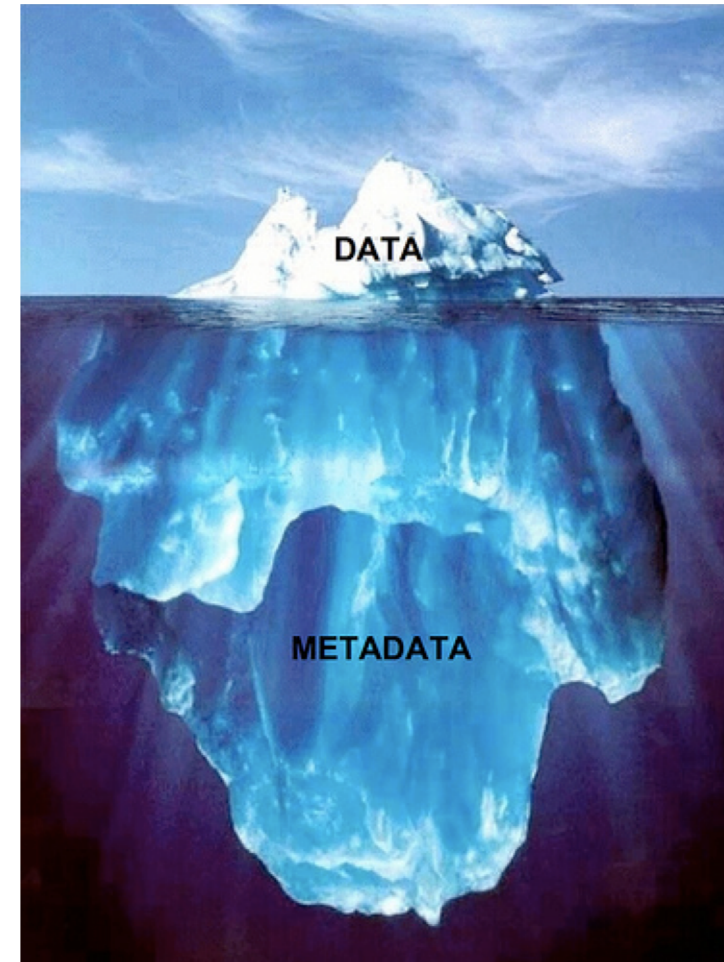
INDEPENDENT REUSE OF DATA

COLLABORATIVE REUSE WITH DATA CREATORS

- Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2018, in review). The Who, What, When, and Why of Reusing Data in Scientific Practice. *Harvard Data Science Review*.
- Borgman, C.L. & Pasquetto, I.V. (2018). *Cochrane Colloquium Edinburgh: Opening keynote*. Edinburgh, UK. Video recording: https://www.youtube.com/watch?v=9PEU_rfkejM&list=PLCo8P5_ppmQgdp0mDMud0CEt8v7DXXgJE

Reuse across place and time

- Reuse by investigator
- Reuse by collaborators
- Reuse by colleagues
- Reuse by unaffiliated others
- Reuse at later times
 - Months
 - Years
 - Decades
 - Centuries



Building the social and technical bridges to enable open data sharing

- Data sharing is a means to many ends
- Time to make tough organizational choices
- Key cases in science, technology, and policy
 - Reusing data within and between communities
 - Sustaining access to data, software, tools, instruments, specimens
 - Sustaining access to data resources beyond grant projects
 - Creating career paths for data stewardship



Christine Borgman



Bernie Boscoe



Peter Darch



Milena Golshan



Irene Pasquetto



Michael Scroggins



Cheryl Thompson



Morgan Wofford

The Data Creators' Advantage

- Data creators have fullest knowledge of scientific goals, context, processing, interpretation
- Data reusers may require contextual information beyond standard documentation
- For foreground reuse, collaboration between data creators and reusers enables
 - Identification of problems of common interest
 - Mutual exchange of expertise
 - Credit through new publications

• Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2018, in review). The Who, What, When, and Why of Reusing Data in Scientific Practice. *Harvard Data Science Review*.

• Borgman, C.L. & Pasquetto, I.V. (2018). *Cochrane Colloquium Edinburgh: Opening keynote*. Edinburgh, UK. Video recording: https://www.youtube.com/watch?v=9PEU_rfkejM&list=PLCo8P5_ppmQgdp0mDMud0CEt8v7DXXgJE