Research Data Infrastructure: A Problem of Governance

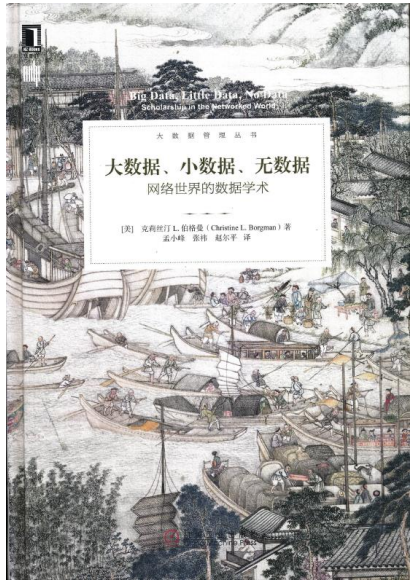
Christine L. Borgman

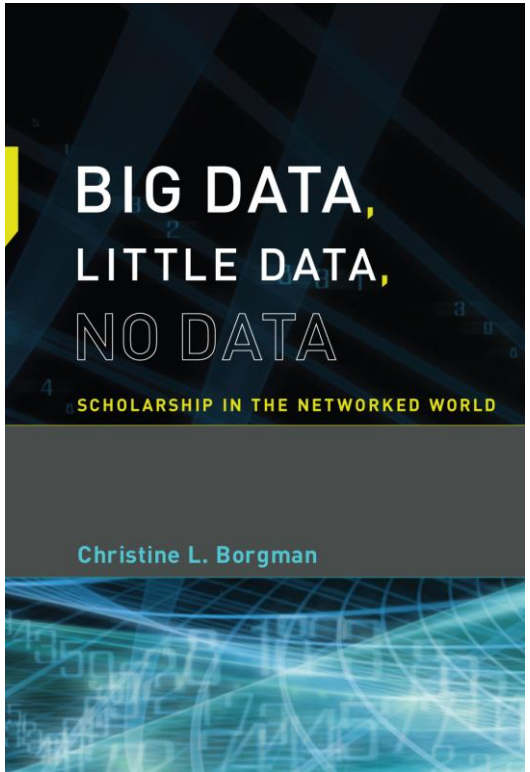
Distinguished Research Professor & Presidential Chair in Information Studies, Emerita

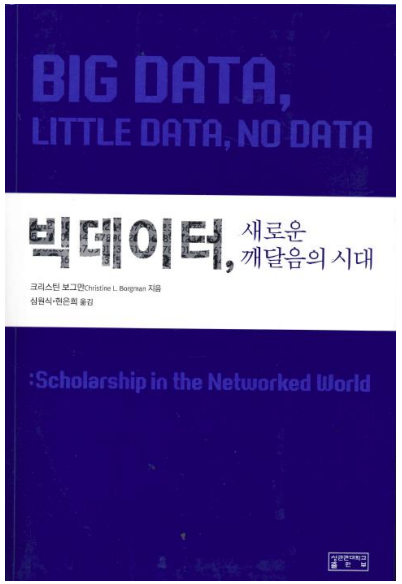
Director, UCLA Center for Knowledge Infrastructures, @scitechprof

Seminario de Estudios sobre el Futuro

El Colegio de Mexico, 7 June 2022, bitly/Futuro-7junio







Background material

Analysis of how universities govern their research data

- Borgman, Christine L., & Bourne, Philip E. (2022). Why it takes a village to manage and share data. *Harvard Data Science Review*, *in press*. http://arxiv.org/abs/2109.01694

Interview study of how universities govern their administrative and research data

- Borgman, Christine L., & Brand, Amy. (2022, under review). Universities are data rich, data poor, and data blind


Royce Hall, UCLA

Open Access / Open Data Policies

- European Research Council
- Research Councils of the UK
- Australian Research Council
- U.S. Federal research policy
- Individual countries, funding agencies, journals, universities

National Science Foundation
WHERE DISCOVERIES BEGIN

JAPAN SCIENCE AND TECHNOLOGY AGENCY

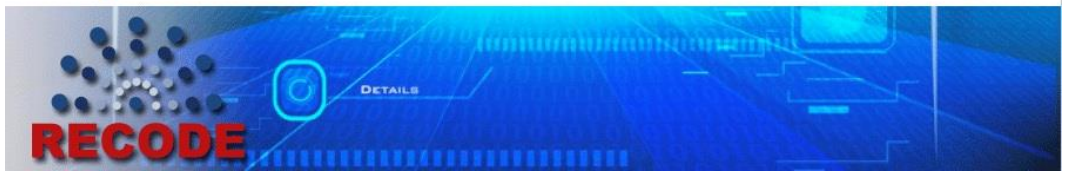
中央研究院
ACADEMIA SINICA

Australian Government
National Health and Medical Research Council

erc European Research Council
Supporting top researchers from anywhere in the world

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Policy RECommendations for Open Access to Research Data in Europe



Sharing Research Data

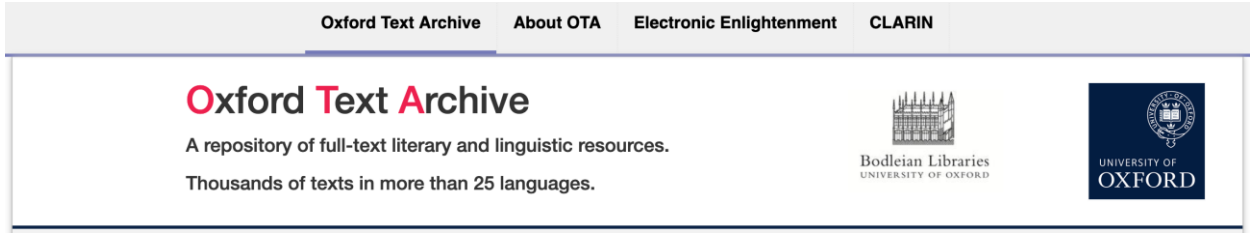
- Link datasets to journal article or publication
- Deposit datasets in a digital data archive
- Publish data documentation
  - Research protocols
  - Codebooks
  - Software
  - Algorithms
- Cite data and software

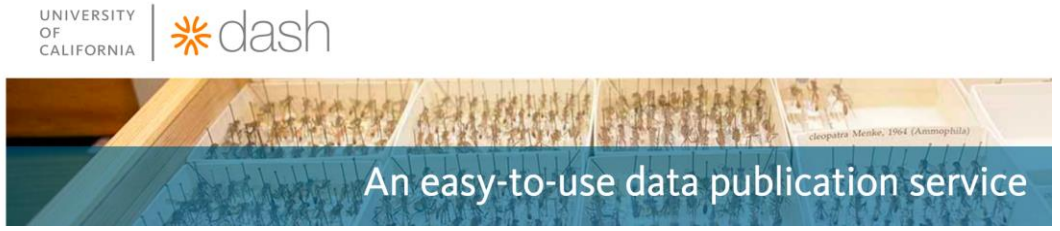
Data Archiving and Networked Services

DANS

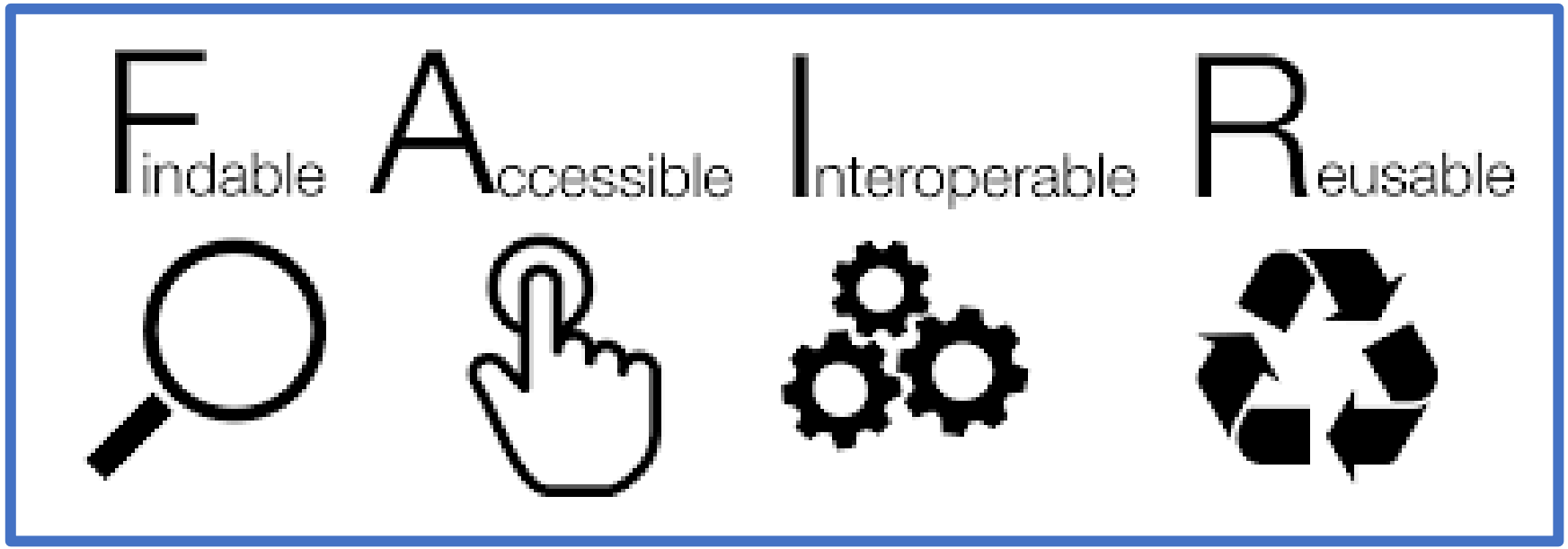
HARVARD Dataverse

ICPSR A PARTNER IN SOCIAL SCIENCE RESEARCH | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH





Data Sharing and Stewardship: The Ideal



5

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

Research data infrastructure

Knowledge infrastructures: “robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds” (Edwards, 2010)

- Policy frameworks
- Scholarly practices
- Technical infrastructures
- Governance models


Royce Hall, UCLA

Edwards, P. N. (2010). *A vast machine: Computer models, climate data, and the politics of global warming*. MIT Press.

Research data infrastructure: Stakeholders

- Research funding agencies
- Individual scientists and scholars
- Academic institutions
  - Academic leadership
  - Research Computing
  - University libraries
  - Schools and departments

Photo by Mihai Surdu on Unsplash

Borgman, C. L., & Bourne, P. E. (2022). Why it takes a village to manage and share data. *Harvard Data Science Review*, *in press*. http://arxiv.org/abs/2109.01694

Stakeholder: Individual scientist

- Roles
  - Principal investigator
  - Collaborator
  - Student, researcher, post-doctoral fellows…
- Responsibilities
  - Data collection and analysis
  - Writing for publication
  - Managing teams
  - Writing grant proposals
  - Managing data, software, technology…

Photo by Mihai Surdu on Unsplash

Borgman, C. L., & Bourne, P. E. (2022). Why it takes a village to manage and share data. *Harvard Data Science Review*, *in press*. http://arxiv.org/abs/2109.01694

Stakeholder: Academic Research Leadership

- Roles
  - Vice president for research
  - Deans and directors…
- Responsibilities
  - Extramural funding
  - Financial management
  - Compliance with regulations
  - Technology transfer
  - Data management
  - Governance of data, privacy, technology…

Photo by Mihai Surdu on Unsplash

Borgman, C. L., & Bourne, P. E. (2022). Why it takes a village to manage and share data. *Harvard Data Science Review*, *in press*. http://arxiv.org/abs/2109.01694

Stakeholder: University Libraries

- Roles
  - Build collections for research and instruction
  - Sustain access to collections
- Responsibilities
  - Maintain knowledge resources
  - Provide physical and online access to resources
  - Promote information literacy
  - Facilitate scholarly communication
  - Steward the scholarly record
  - Construct and maintain data repositories…

Photo by Mihai Surdu on Unsplash

Borgman, C. L., & Bourne, P. E. (2022). Why it takes a village to manage and share data. *Harvard Data Science Review*, *in press*. http://arxiv.org/abs/2109.01694

Research data interdependencies

- What data to share
- Data, context, and credit
- Data and discovery
- Data assets as research methods
- Intellectual property in data
- Data science initiatives
- Thinking globally, acting locally


Royce Hall, UCLA

Borgman, C. L., & Bourne, P. E. (2022). Why it takes a village to manage and share data. *Harvard Data Science Review*, *in press*. http://arxiv.org/abs/2109.01694

Interdependencies: What data to share

Data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship.

C.L. Borgman (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press

http://www.genome.gov/dmd/img.cfm?node=Photos/Graphics&id=85327

National Institutes of Health Data Sharing Policy 2023

Section II. Definitions

For the purposes of the DMS Policy, terms are defined as follows:

SCIENTIFIC DATA	*The recorded factual material commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications. Scientific data do not include laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens.*
DATA MANAGEMENT	*The process of validating, organizing, protecting, maintaining, and processing scientific data to ensure the accessibility, reliability, and quality of the scientific data for its users.*
DATA SHARING	*The act of making scientific data available for use by others (e.g., the larger research community, institutions, the broader public), for example, via an established repository.*
METADATA	*Data that provide additional information intended to make scientific data interpretable and reusable (e.g., date, independent sample and variable construction and description, methodology, data provenance, data transformations, any intermediate or descriptive observational variables).*
DATA MANAGEMENT AND SHARING PLAN (PLAN)	*A plan describing the data management, preservation, and sharing of scientific data and accompanying metadata.*

Interdependencies: Data, Context, and Credit

- Publications
  - Independent units
  - Authorship is negotiated
- Data
  - Compound objects
  - Ownership is rarely clear
  - Attribution
    - Long term responsibility: Investigators
    - Expertise for interpretation: Data collectors and analysts
- Representation and interpretation

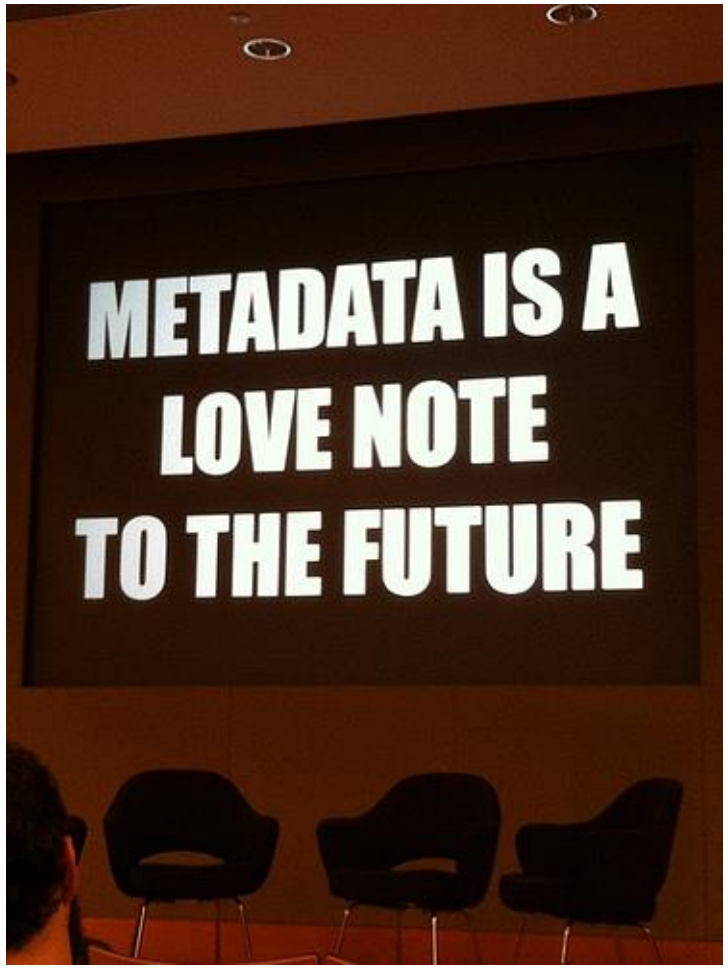
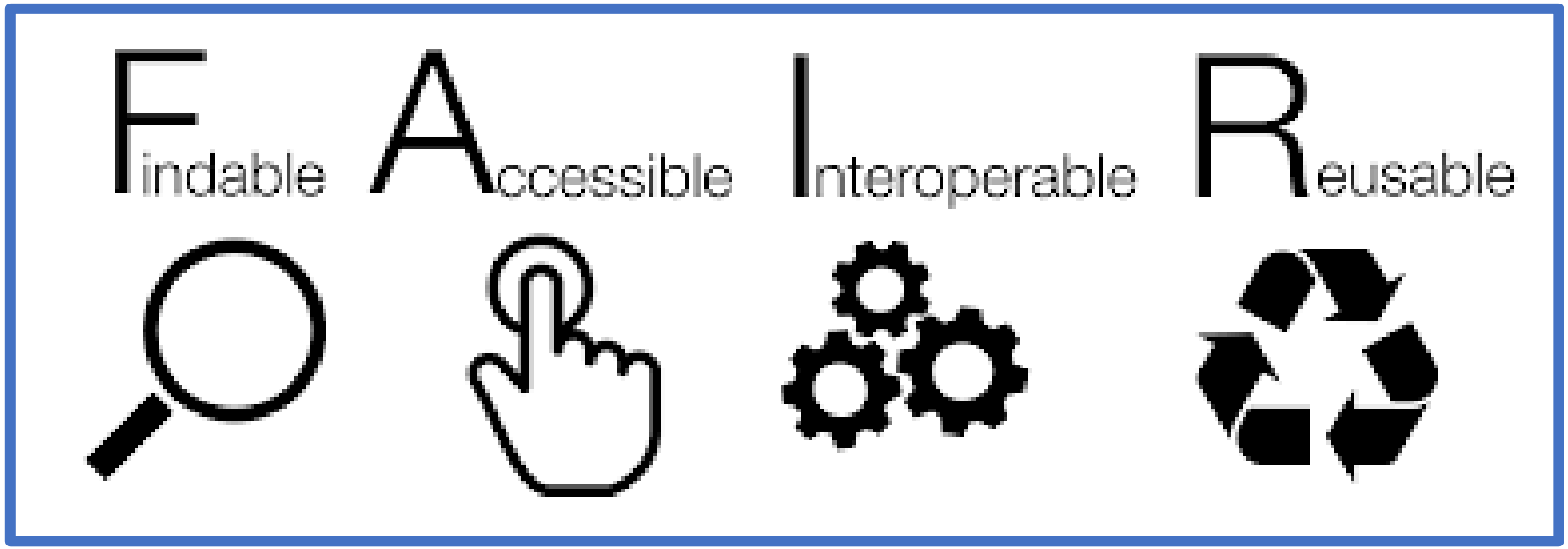


Photo by @kissane; presentation by Jason Scott (@textfiles)

Borgman (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press
Borgman (2016). Data Citation as a Bibliometric Oxymoron. *Theories of Informetrics and Scholarly Communication*. https://doi.org/10.1515/9783110308464-008

Data Sharing and Stewardship: The Ideal

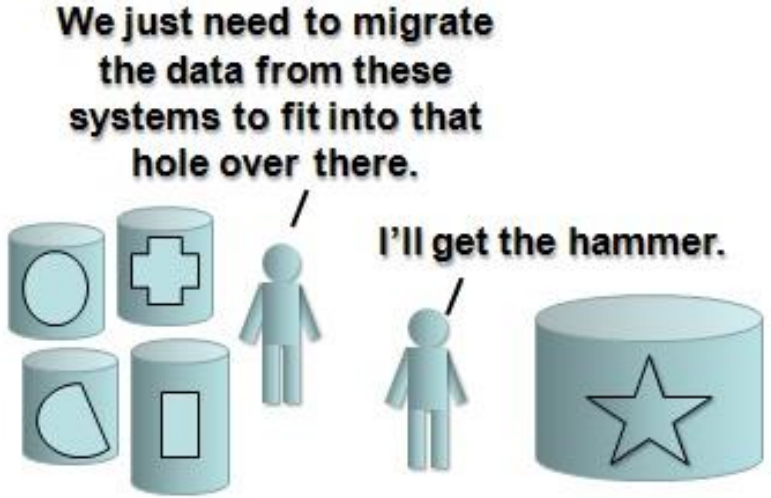


15

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

Data Stewardship: The Reality

http://www.information-age.com/cloud-computing-pharmaceutical-industry-123462676/



http://www.datamartist.com/data-migration-part-1-introduction-to-the-data-migration-delema

Mount Wilson Solar Observatory, 2017

Getty Research Institute

http://gsa.rice.edu/

Graduate students

https://med.nyu.edu/our-community/life-nyu-school-medicine/life-postdoc

Post-doctoral fellows

Interdependencies: Thinking globally, acting locally

- Research data
  - Value is international
  - Funding is national
- Knowledge infrastructure
  - Local institutions and campuses
  - National funding
  - International coordination
- Conflicting governance
  - Data sharing policies
  - Privacy laws, transborder data flows
  - Intellectual property…

Borgman, C. L., & Bourne, P. E. (2022). Why it takes a village to manage and share data. *Harvard Data Science Review*, *in press*. http://arxiv.org/abs/2109.01694

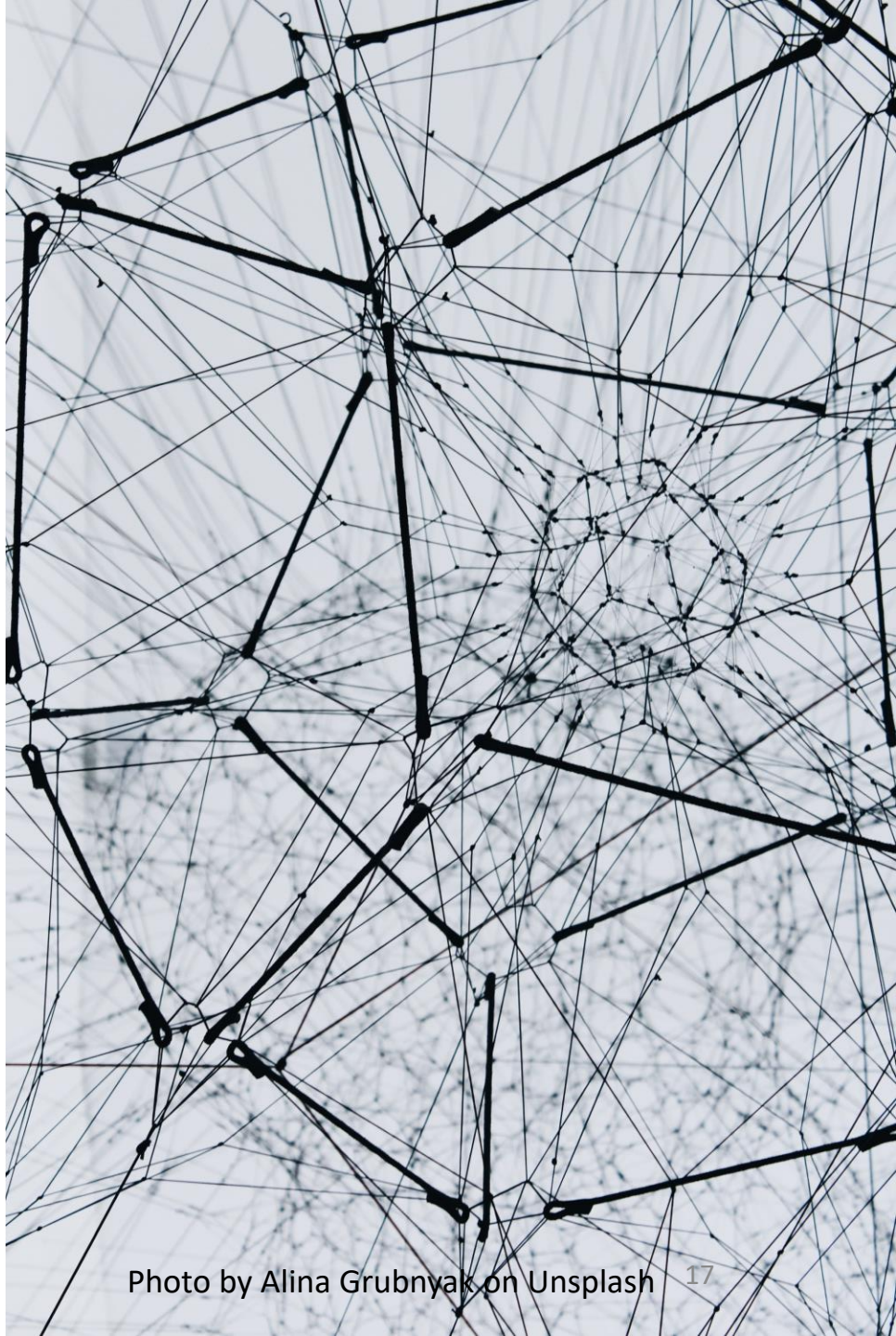
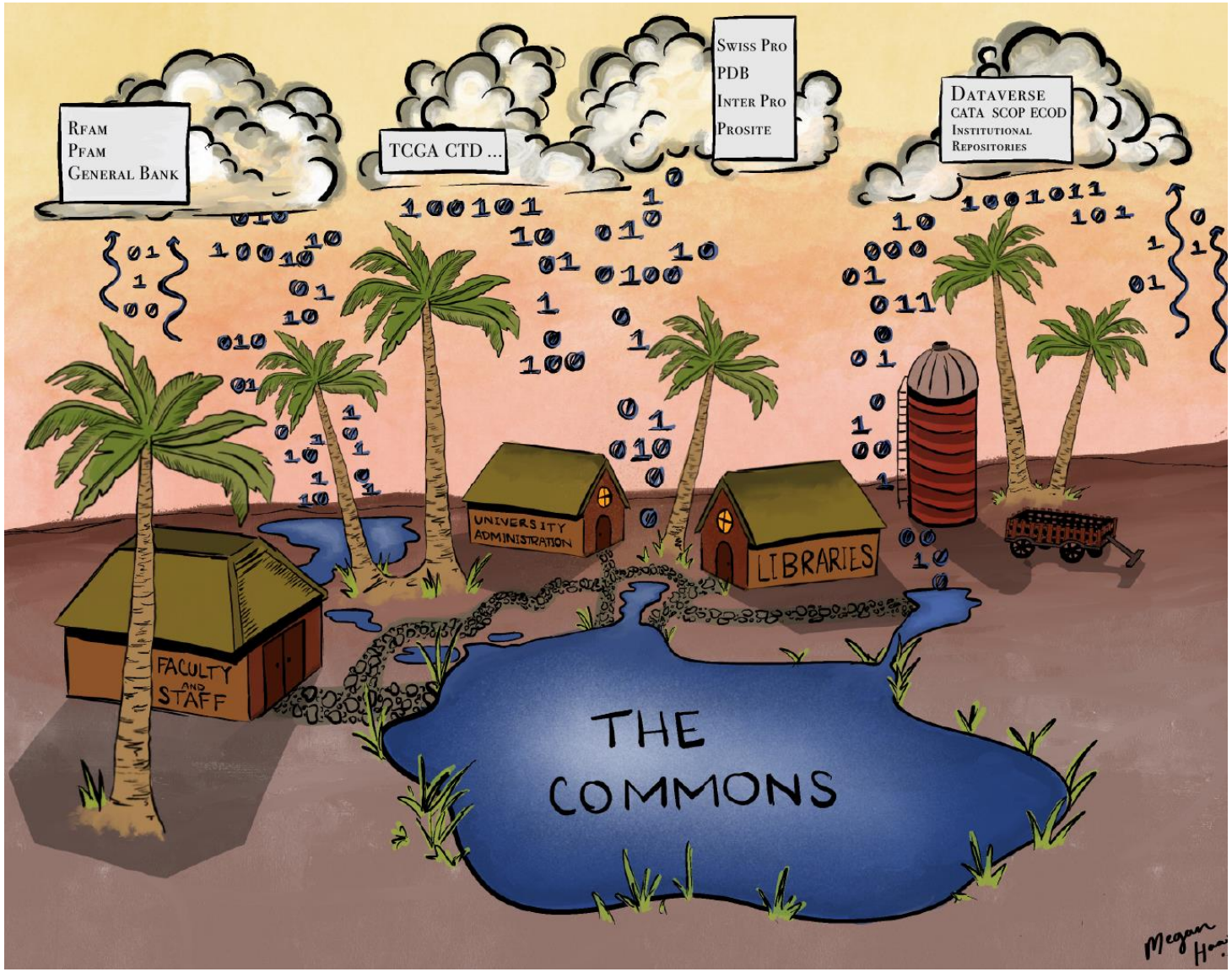


Photo by Alina Grubnyak on Unsplash

Research Data Governance:

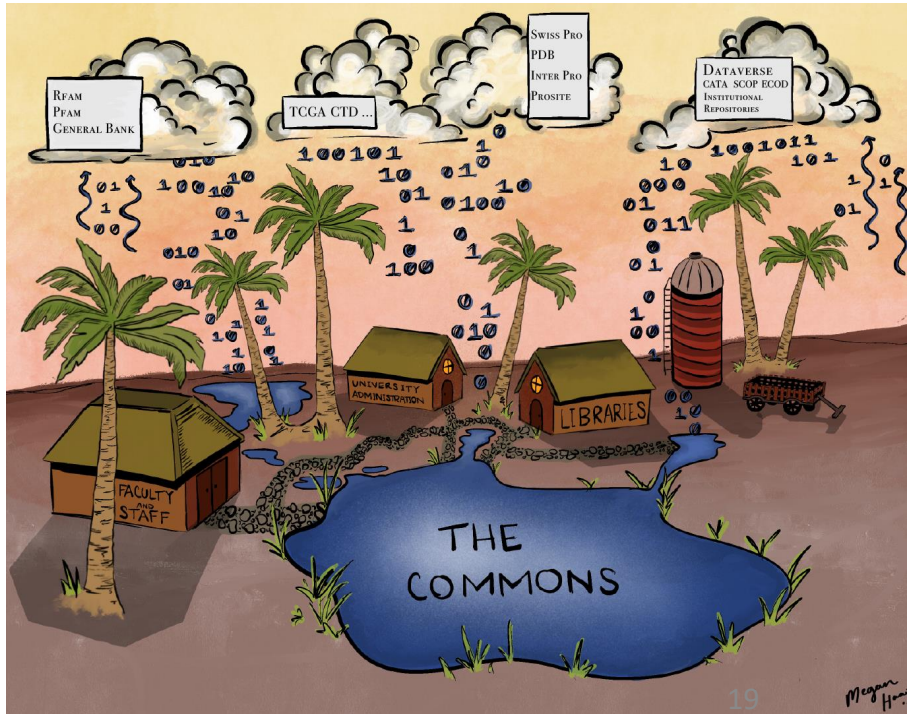
Building the Village

Borgman, C. L., & Bourne, P. E. (2022). Why it takes a village to manage and share data. *Harvard Data Science Review.* Illustration by Megan Haas



Governance: Building the Village

- Data sharing is a ‘collective action problem’
- Holistic approaches to sharing infrastructure
  - Distribute responsibility among stakeholders
  - Invest in data management expertise
  - Reframe goals in collective terms
- Fund the commons
  - Public support for data repositories
  - International exchange of best practices
- Invest in sustainable strategies



Borgman, C. L., & Bourne, P. E. (2022). Why it takes a village to manage and share data. *Harvard Data Science Review.* Illustration by Megan Haas

Discussion questions

- Who should be involved in governing research data within a university?
- Who should be involved in governing research data nationally? Internationally?
- What governance criteria should apply to
  - Releasing research data
  - Using others' research data
  - Stewarding research data
  - Sustaining infrastructure for research data

Bodleian Library, University of Oxford; Borgman photo 2004

Further reading

- Aspesi, C., & Brand, A. (2020). In pursuit of open science, open access is not enough. *Science*, *368*(6491), 574–577. https://doi.org/10.1126/science.aba3763
- Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. MIT Press.
- 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
- Borgman, C. L. (2020). Whose text, whose mining, and to whose benefit? *Quantitative Science Studies*, *1*(3), 993–1000. https://doi.org/10.1162/qss_a_00053
- Borgman, C. L., & Bourne, P. E. (2022). Why it takes a village to manage and share data. *Harvard Data Science Review*, *in press*. http://arxiv.org/abs/2109.01694
- Borgman, C. L., & Brand, A. (2022, in review). Universities are data rich, data poor, and data blind.
- Bourne, P. E., et.al (2022). A Call to US Funders and Policy Makers – Establish the Open Research Commons. *Science*, *in press; et al; placeholder ref*.
- Brand, A. (2022, April 8). *Open access loses when publishers are vilified*. Times Higher Education (THE). https://www.timeshighereducation.com/opinion/open-access-loses-when-publishers-are-vilified
- Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2019). Uses and Reuses of Scientific Data: The Data Creators' Advantage. *Harvard Data Science Review*, *1*(2). https://doi.org/10.1162/99608f92.fc14bf2d