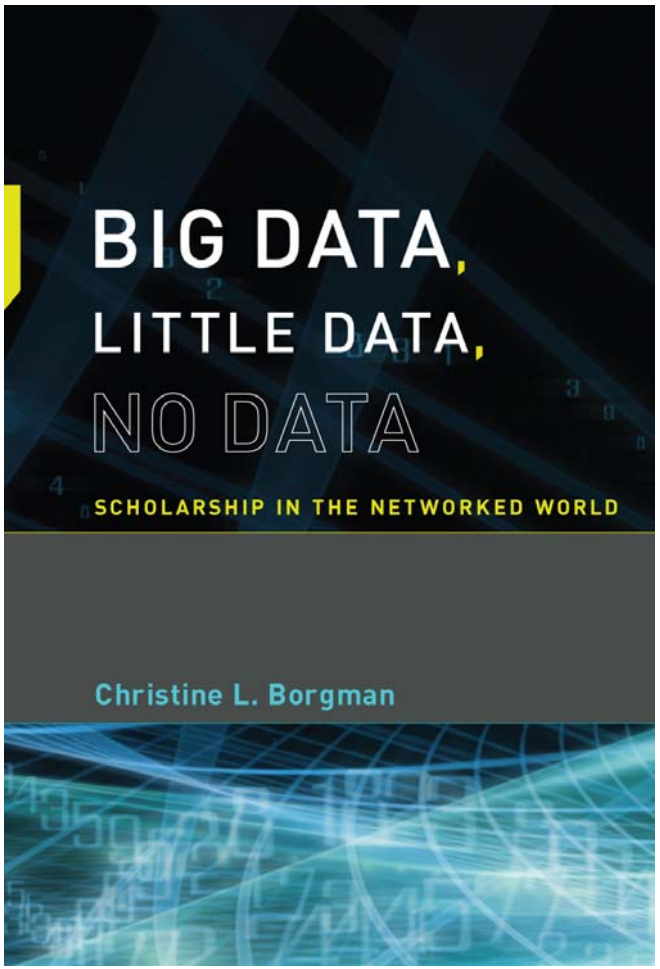
When and Why Should Research Data be Sustained?

Christine L. Borgman
Distinguished Professor
& Presidential Chair in Information Studies
University of California, Los Angeles
Christine.Borgman@UCLA.edu

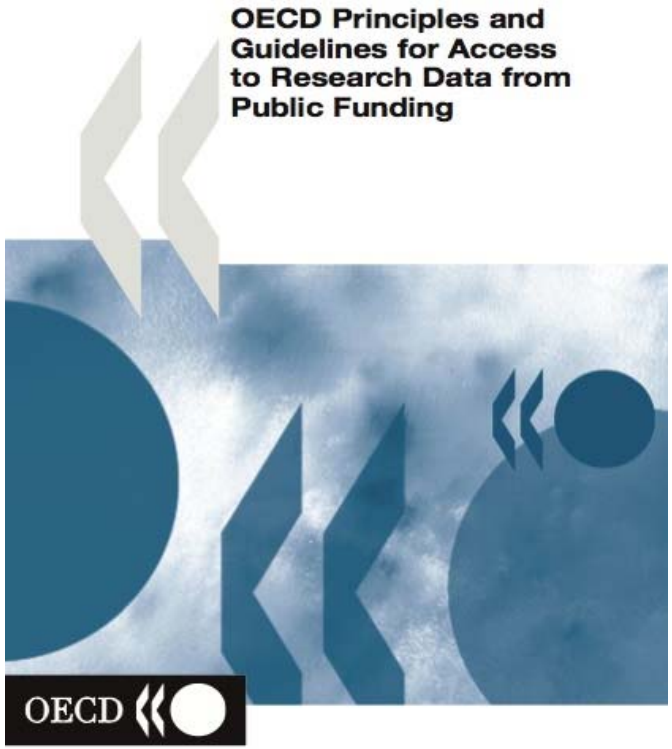
Center for Knowledge Infrastructures
https://knowledgeinfrastructures.gseis.ucla.edu/

National Science Foundation Workshop
Cyberinfrastructure for Large Facilities
December 1-2, 2015



Open Data: OECD criteria

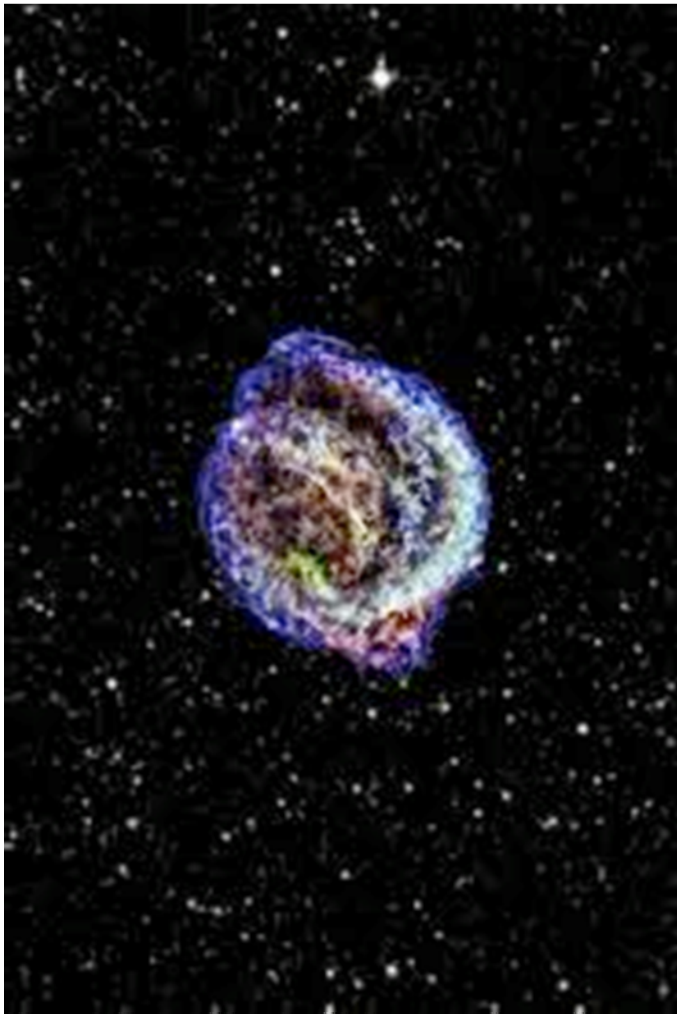
- Openness
- flexibility
- transparency
- legal conformity
- protection of intellectual property
- formal responsibility
- professionalism
- interoperability
- quality
- security
- efficiency
- accountability
- sustainability



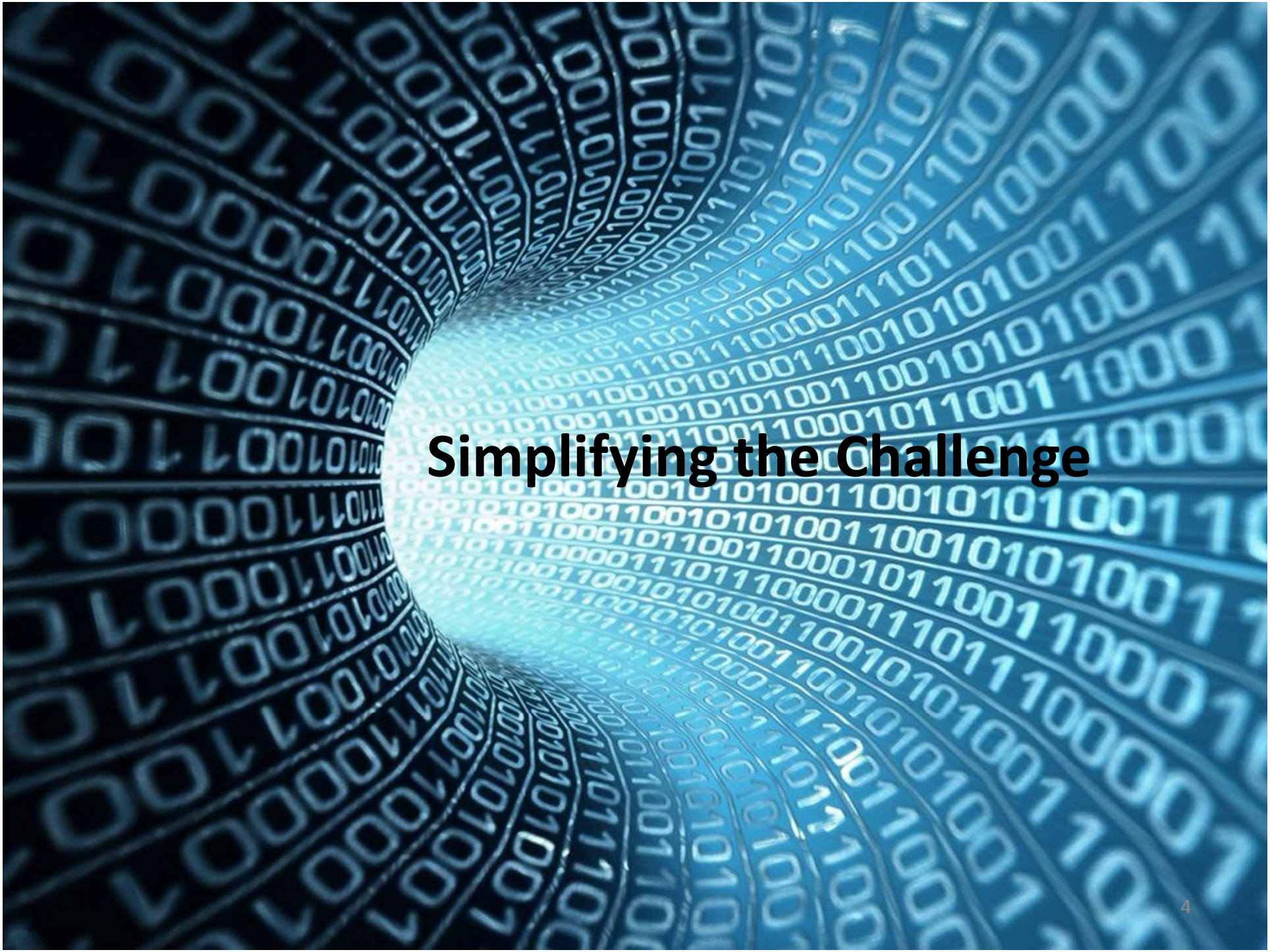
Organization for Economic Cooperation and Development (2007)
http://www.oecd.org/science/sci-tech/38500813.pdf

Why sustain access to data?

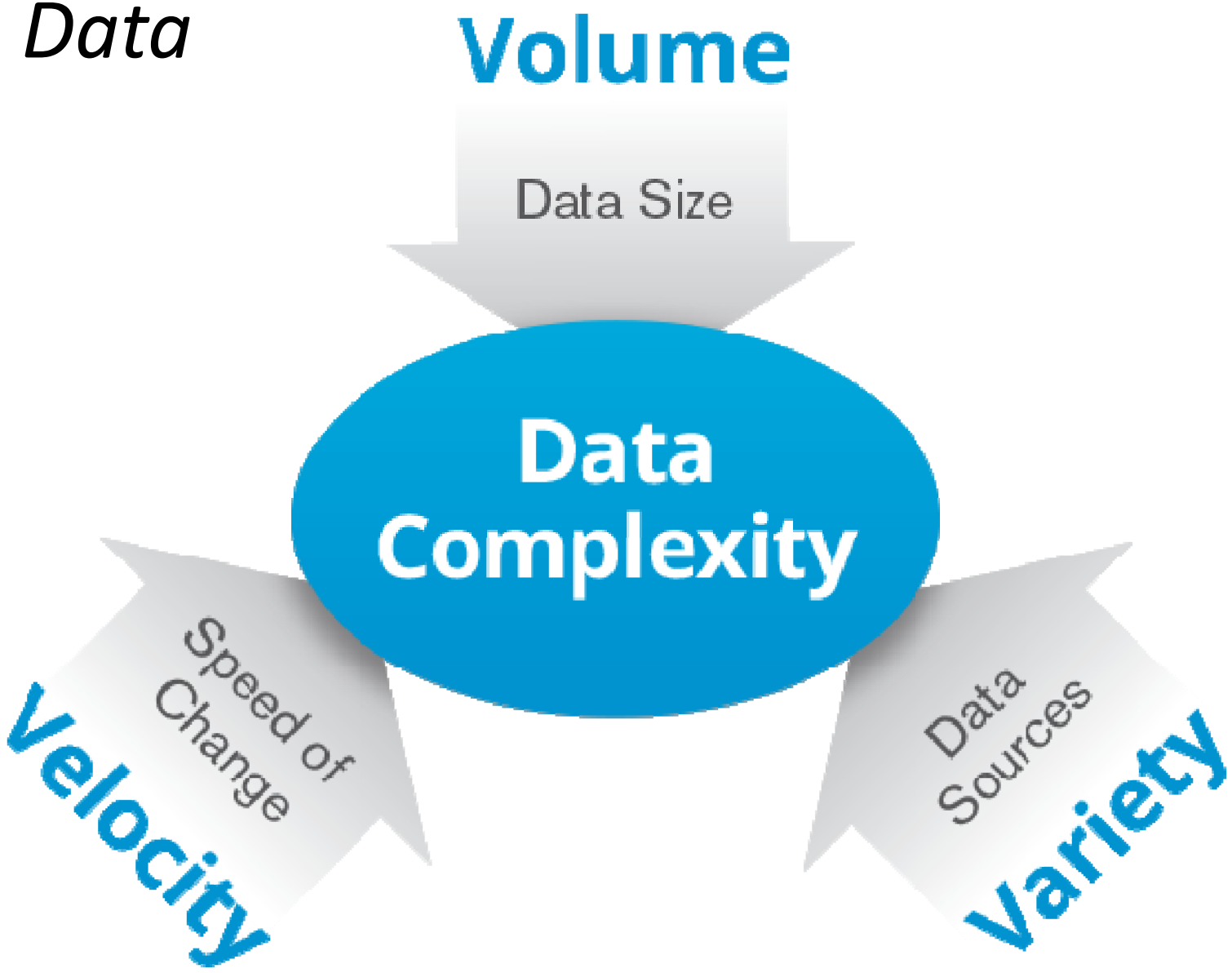
- Purposes
  - Record of observations
  - Reference
  - Reproducibility of research
  - Aggregation from multiple sources
- Users
  - Investigator
  - Collaborators
  - Unaffiliated or unknown others
- Time frame
  - Months
  - Years
  - Decades
  - Centuries



http://chandra.harvard.edu/photo/2013/kepler/kepler_525.jpg

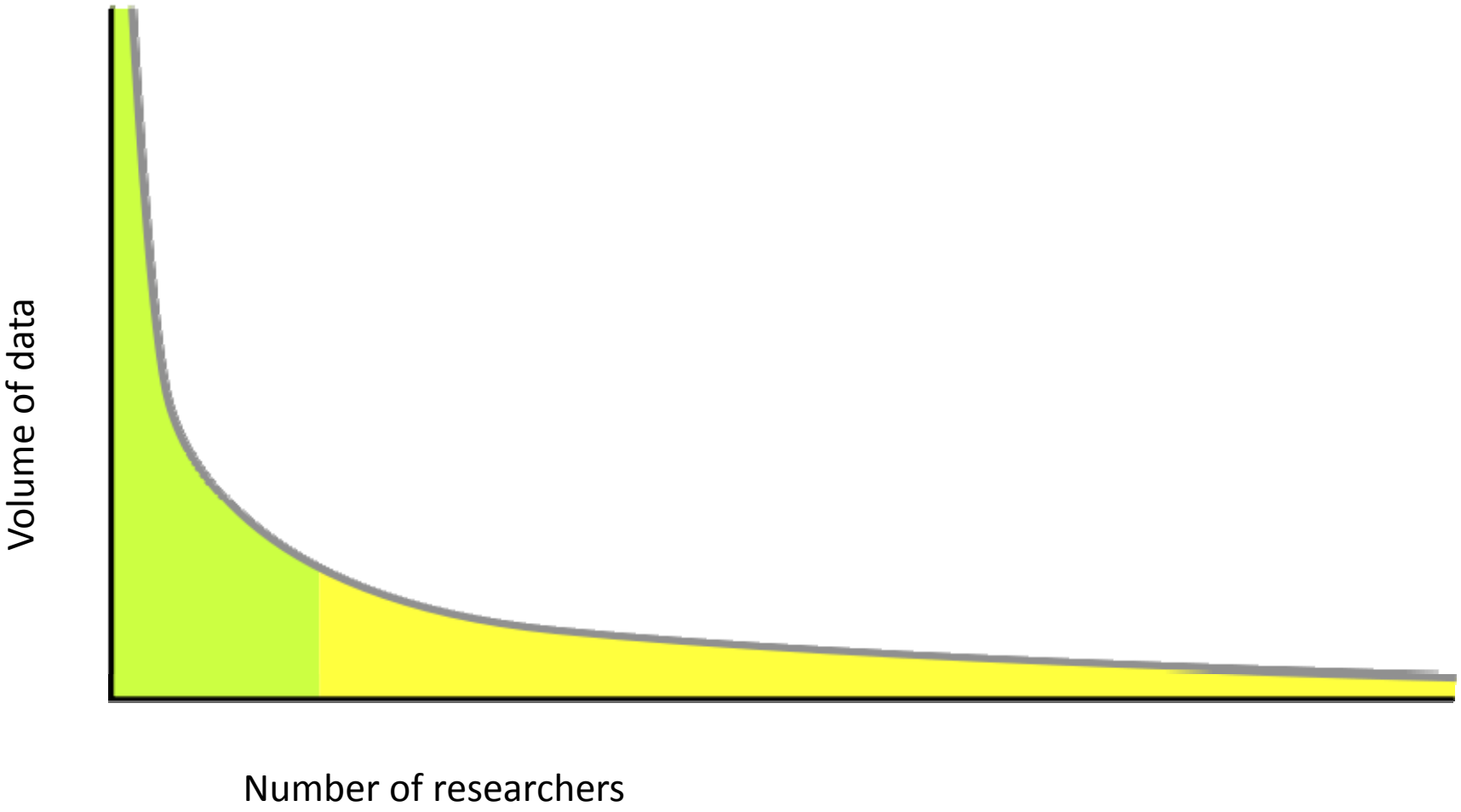

Simplifying the Challenge

Big Data



http://www.datameer.com/product/hadoop.html

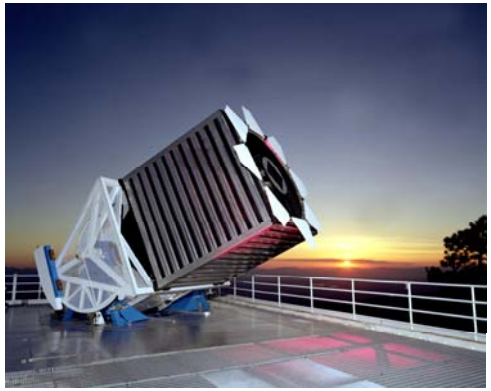
Long tail of data



Slide: The Institute for Empowering Long Tail Research

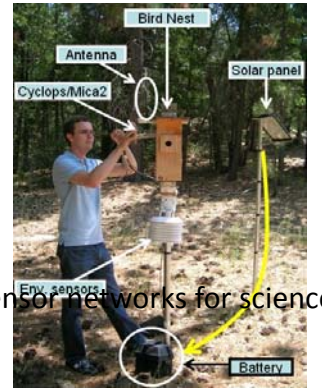
Big Science <–> Little Science

- Large instruments
- High cost
- Long duration
- Many collaborators
- Distributed work
- Centralized data collection



Sloan Digital Sky Survey

- Small instruments
- Low cost
- Short duration
- Small teams
- Local work
- Decentralized data collection



Sensor networks for science

How to sustain data?

- Identify the form and content
- Identify related objects
- Interpret
- Evaluate
- Open
- Read
- Compute upon
- Reuse
- Combine
- Describe
- Annotate…

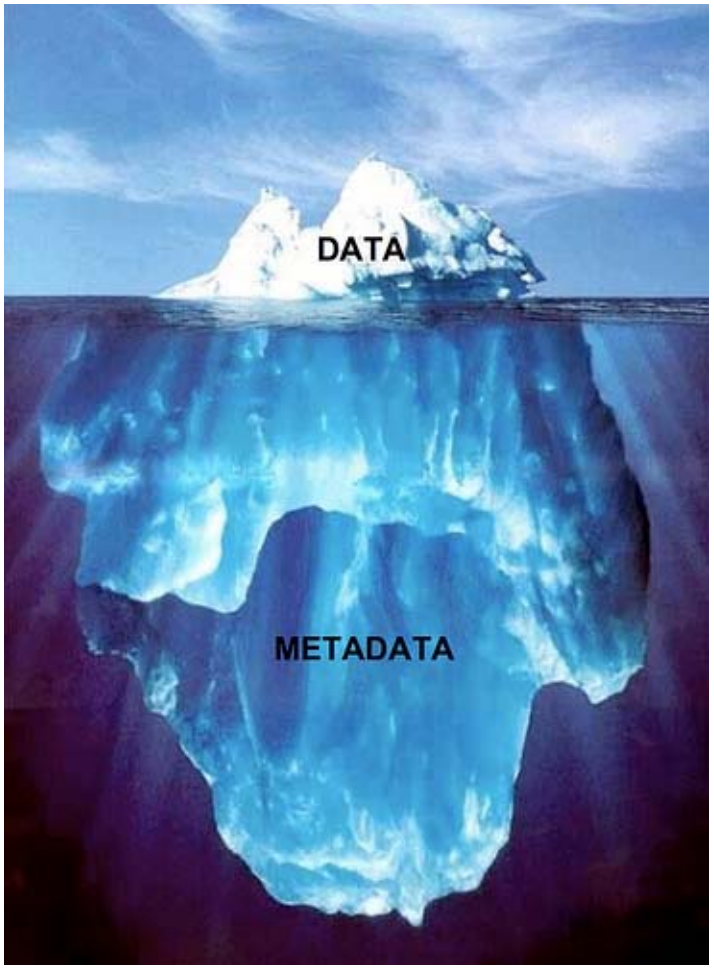
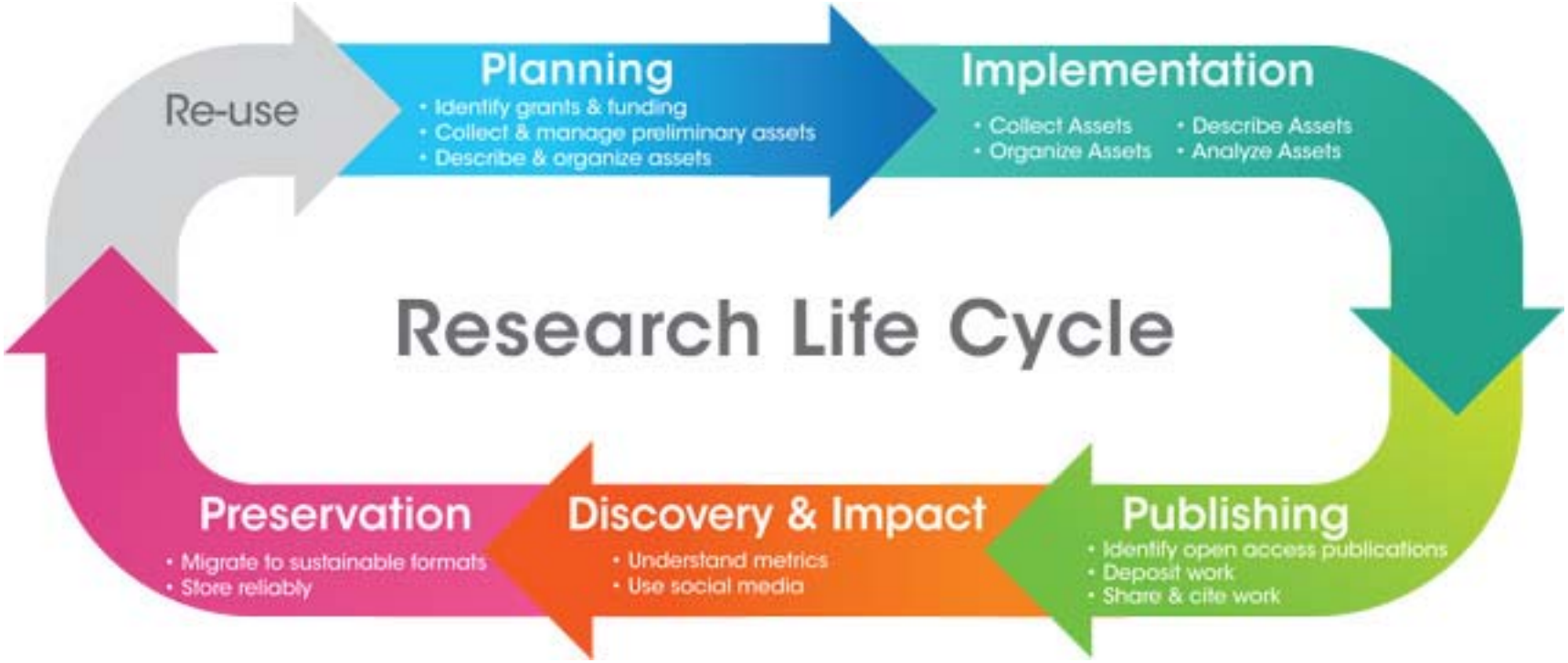


Image from Soumitri Varadarajan blog. Iceberg image © Ralph A. Clevenger. Flickr photo

When to invest in data?



http://www.lib.uci.edu/dss/images/lifecycle.jpg

When to invest in data?

http://www.finance.umich.edu/programs

Economics of the Knowledge Commons

		Subtractability / Rivalry	
		Low	High
Exclusion	Difficult	**Public Goods** General knowledge Public domain data	**Common-pool resources** Libraries Data archives
	Easy	**Toll or Club Goods** Subscription journals Subscription data	**Private Goods** Printed books Raw or competitive data

Adapted from C. Hess & E. Ostrom (Eds.), *Understanding knowledge as a commons: From theory to practice*. MIT Press.

http://knowledgeinfrastructures.org

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

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