Data Management for Collaboration, Access, and Interoperability

Karl Benedict

Plato L. Smith II

University of New Mexico

CLIR Webinar – February 11, 2015

Contents

- Introduction

- Research and Data Lifecycle Models

- Data Management Considerations

- Data Interoperability and Linked Open Data

Introduction

Outline

- Data Management Planning – Foundation Principles
  - Context – Data Management Requirements
  - Relationships Between the Researcher and Data Lifecycles Models – Part 1
  - Relationships Between the Researcher and Data Lifecycles Models – Part 2
  - Data Interoperability and Linked Open Data

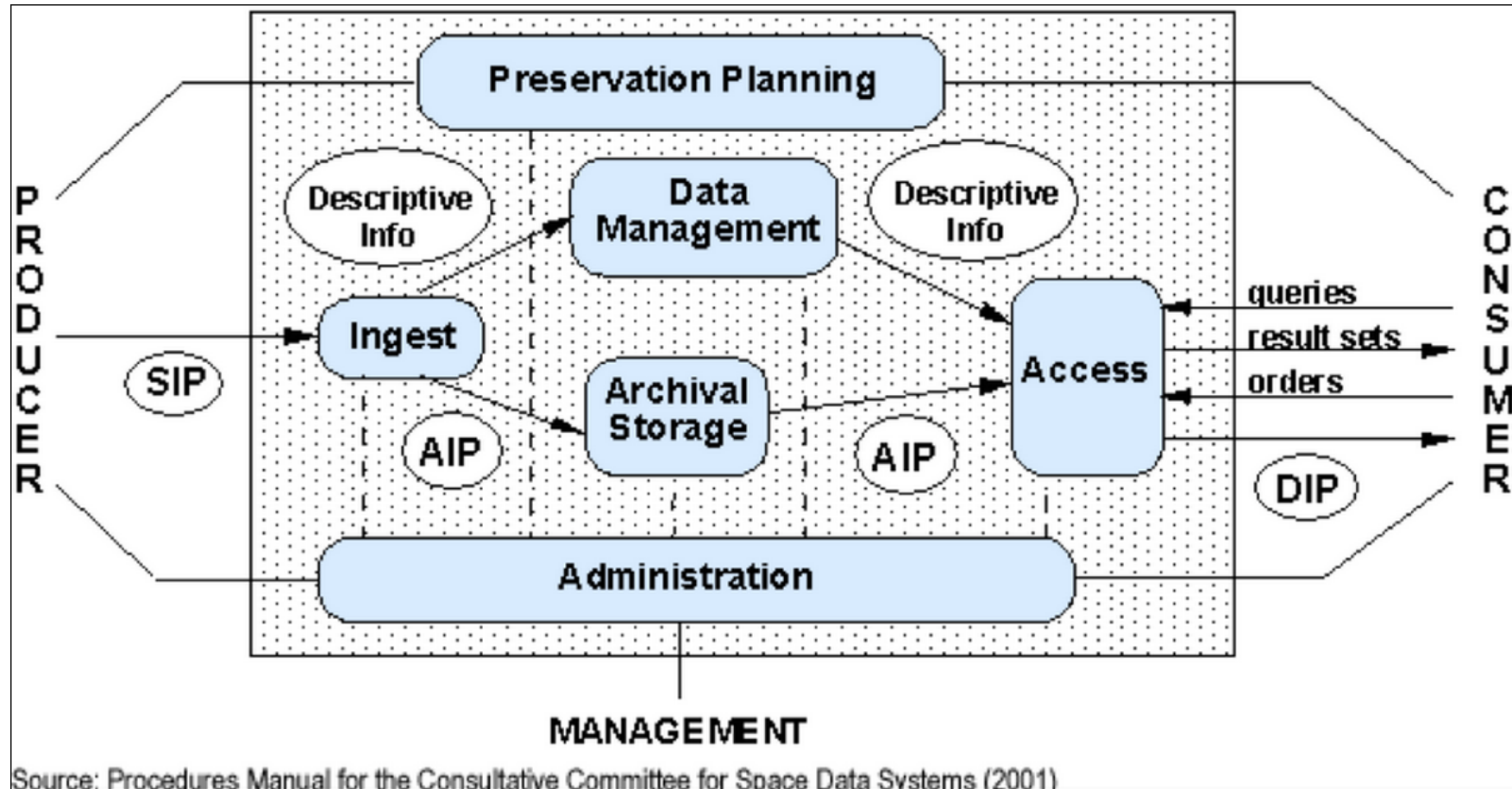# Context – Data Management Requirements

- Data Management Plans
- Data Sharing Requirements
- Institutional Review Board (IRB) Protocols
- Interdisciplinary Collaborative Research
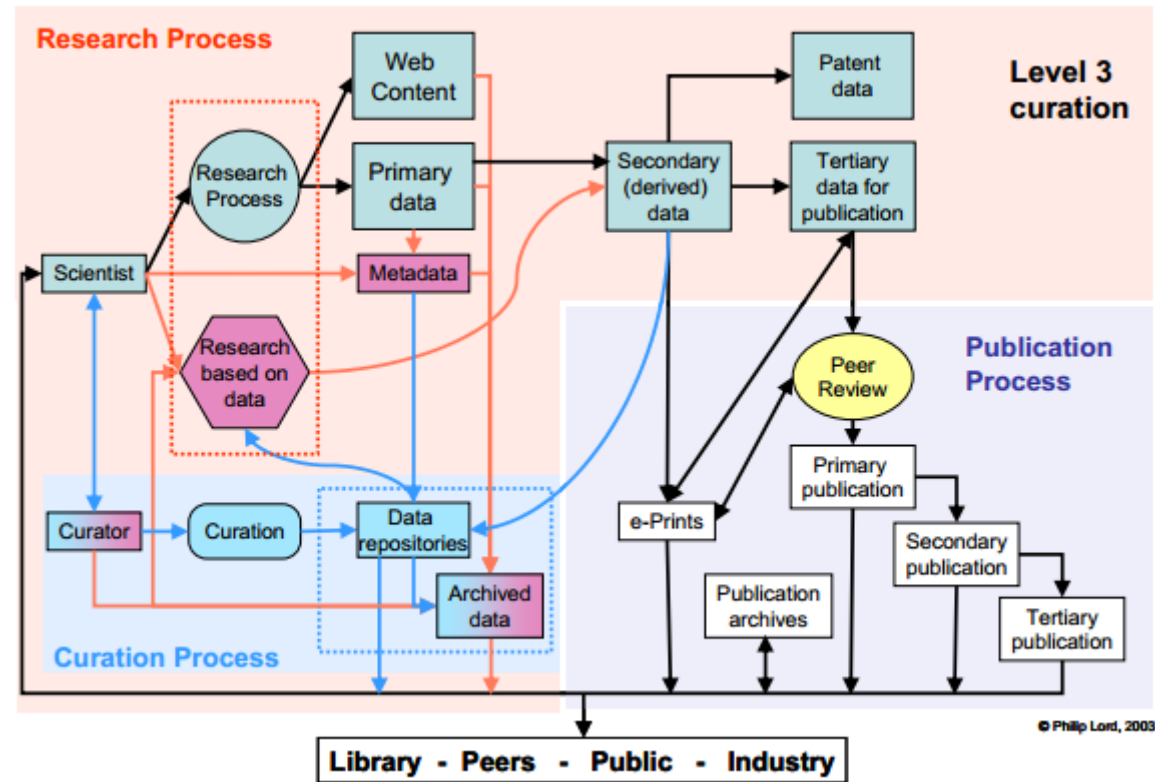- Data Intensive Research

Research and Data Lifecycle Models

Relationship Between the Researcher and Data Lifecycles Models – Part 1 (circa 2001 to 2011)

Source: Procedures Manual for the Consultative Committee for Space Data Systems (2001)

- 2001 CCSDS 650.0-R-2 (Red Book)

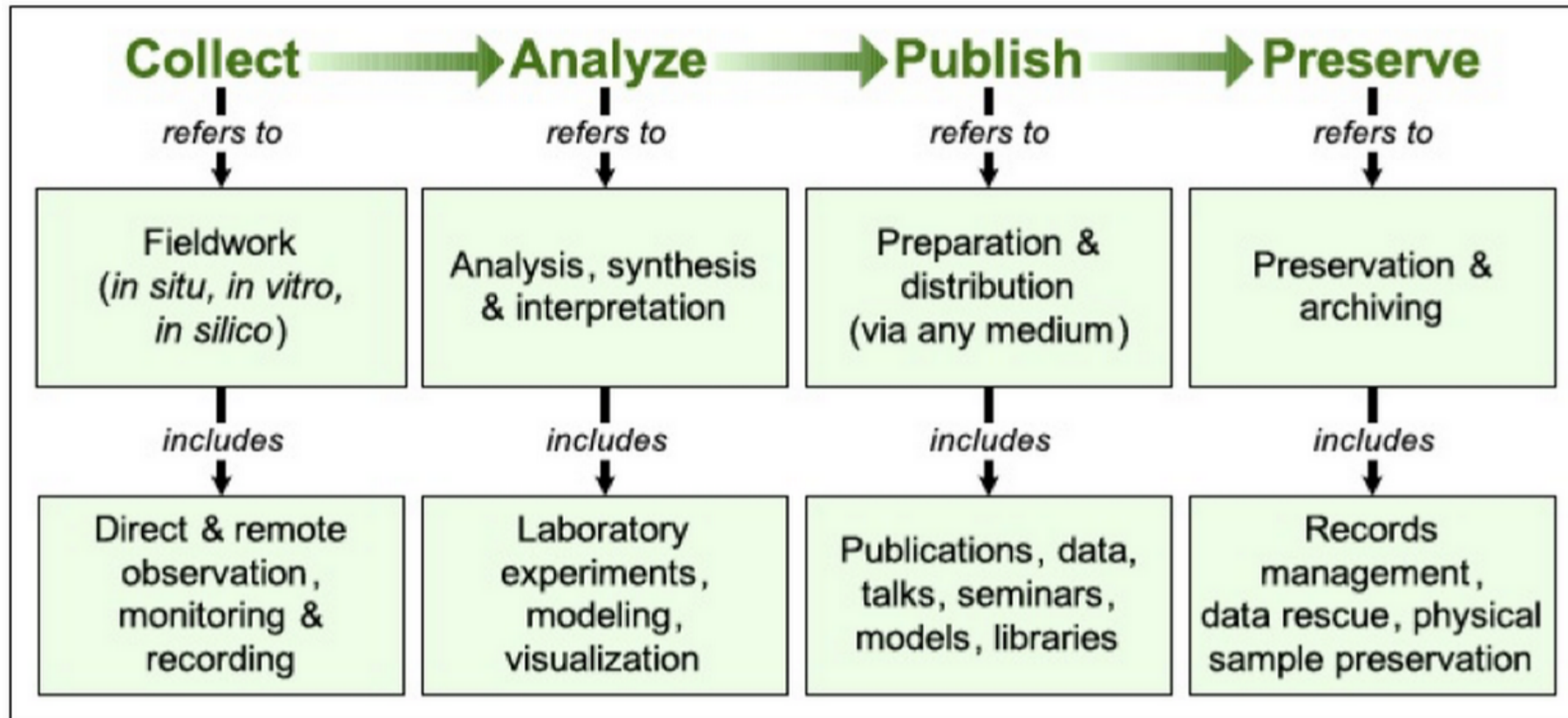# Relationship Between the Researcher and Data Lifecycles Models – Part 1 (circa 2001 to 2011)
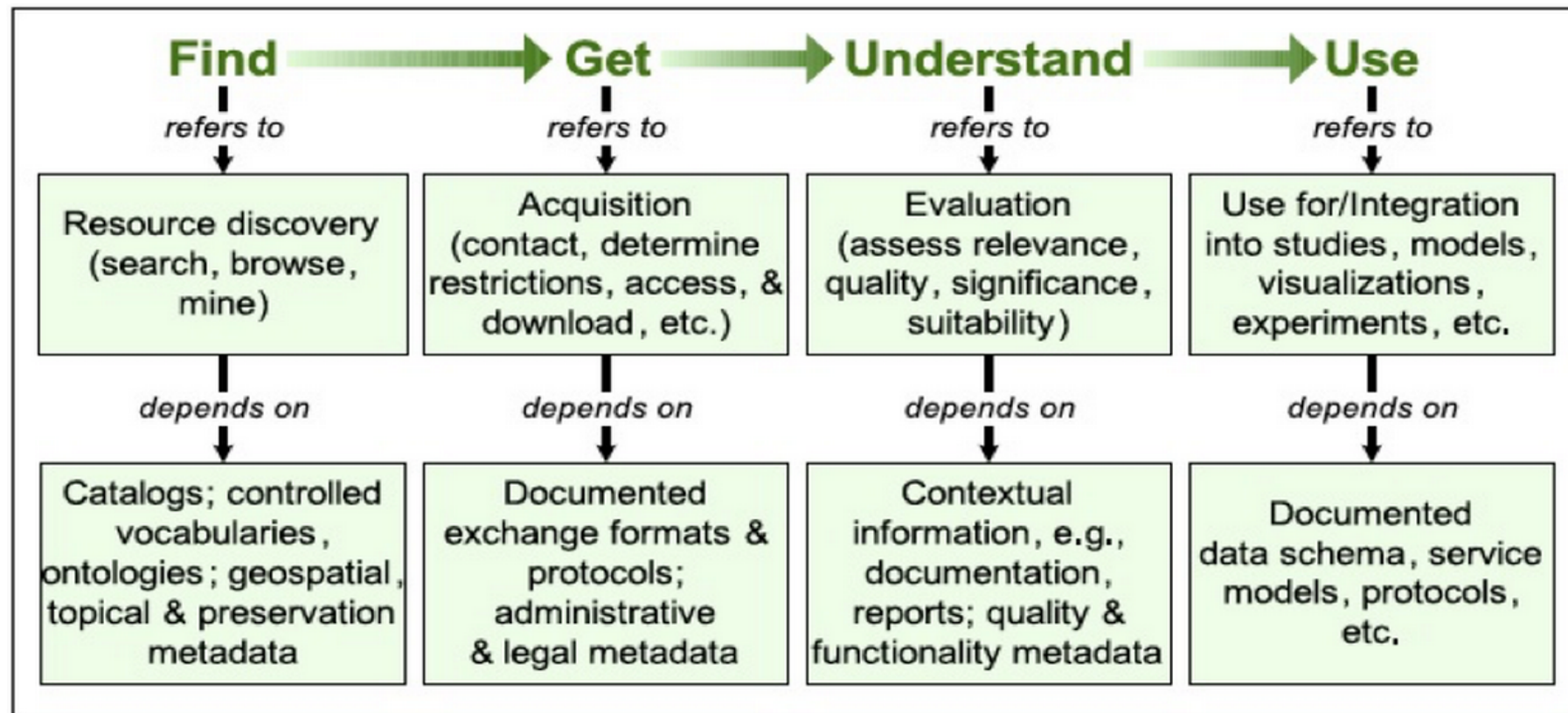

- 2003 e-Science Curation Report (Lord & Macdonald)

Relationship Between the Researcher and Data Lifecycles Models – Part 1 (circa 2001 to 2011)
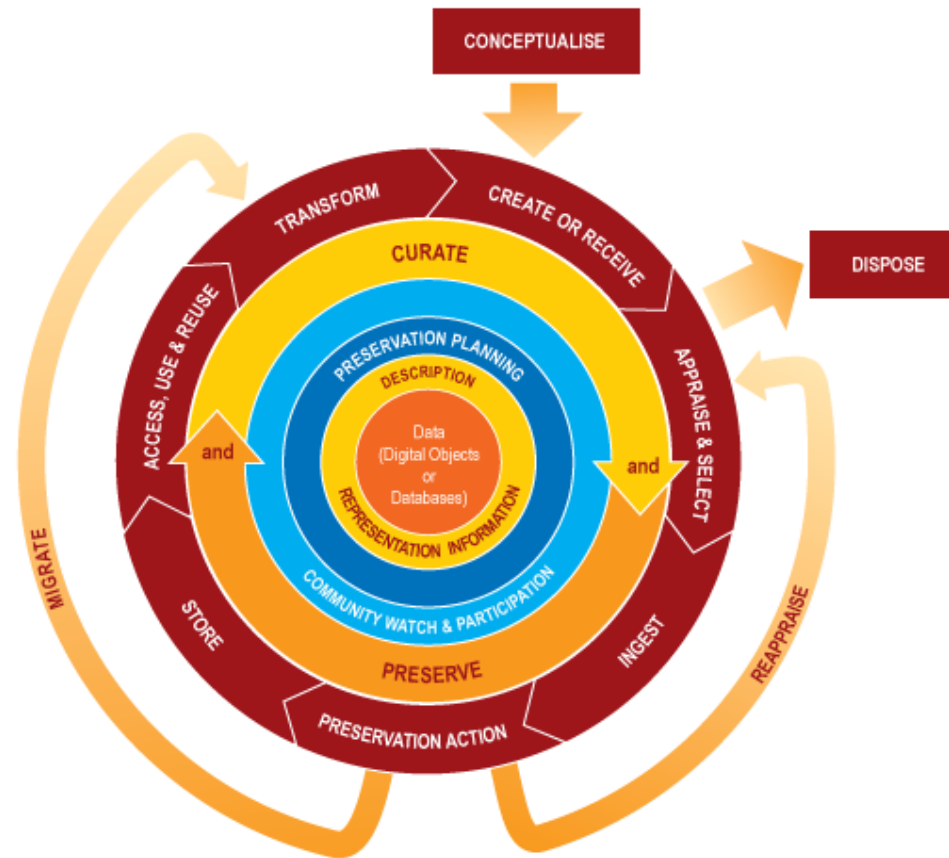
Producer Perspective

- 2006 USGS (Govoni and Gunther)

Relationship Between the Researcher and Data Lifecycles Models – Part 1 (circa 2001 to 2011)

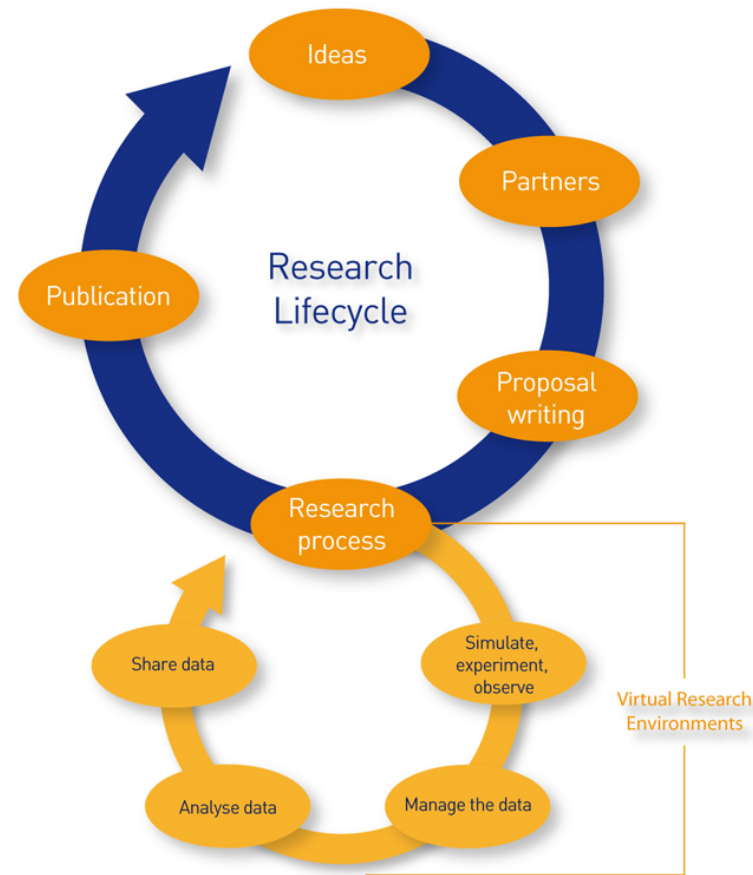**Consumer Perspective**

- 2006 USGS (Govoni and Gunther)

Relationship Between the Researcher and Data Lifecycles Models – Part 1 (circa 2001 to 2011)

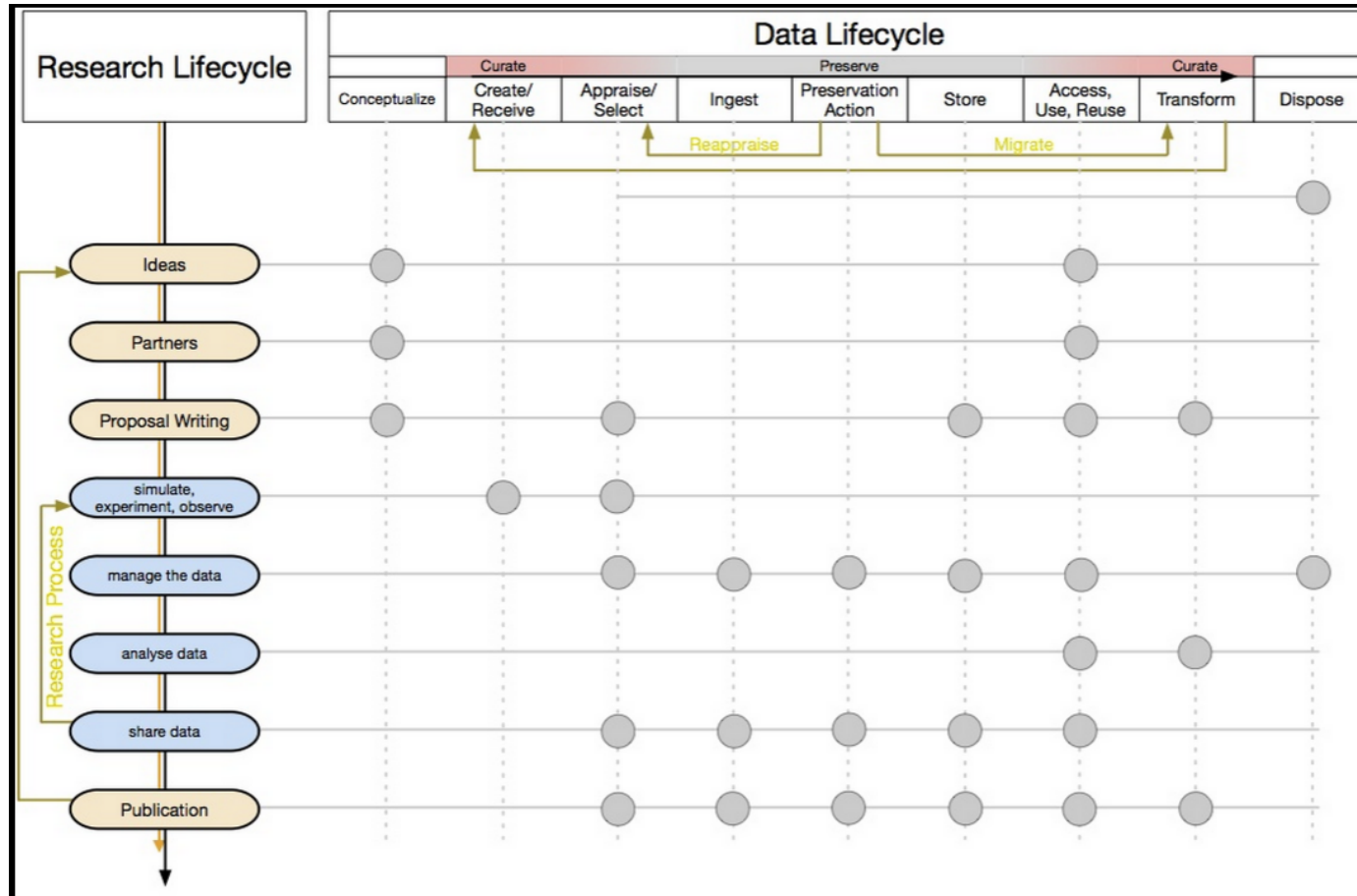
- 2007/2015 DCC Curation Lifecycle Model

Relationship Between the Researcher and Data Lifecycles Models – Part 2 (circa 2013 to Present)

- [2013 JISC Research Lifecycle Diagram](#)

Relationship Between the Researcher and Data Lifecycles Models – Part 2 (circa 2013 to Present)

- 2013/2014 Mapping Research Lifecycle to Data Lifecycle (Benedict)

Data Management Considerations

Some Definitions

- Data
- Data Curation
- Documentation (Metadata)
- Open Access
  - Consent to Share
- Embargo
  - Consent to Restrict
- License
- Data Repositories
- Long-term preservation
  - Standards

Bruegel's Tower of Babel

Some Recommendations

- What do you need to know?
  - *Four Kinds of Expertise
    - Domain (Subject)
    - Analytical
    - Data Management
    - Project Management
  - Professional Development and Training
- Data Assessment
  - Organization
  - Structure/Content
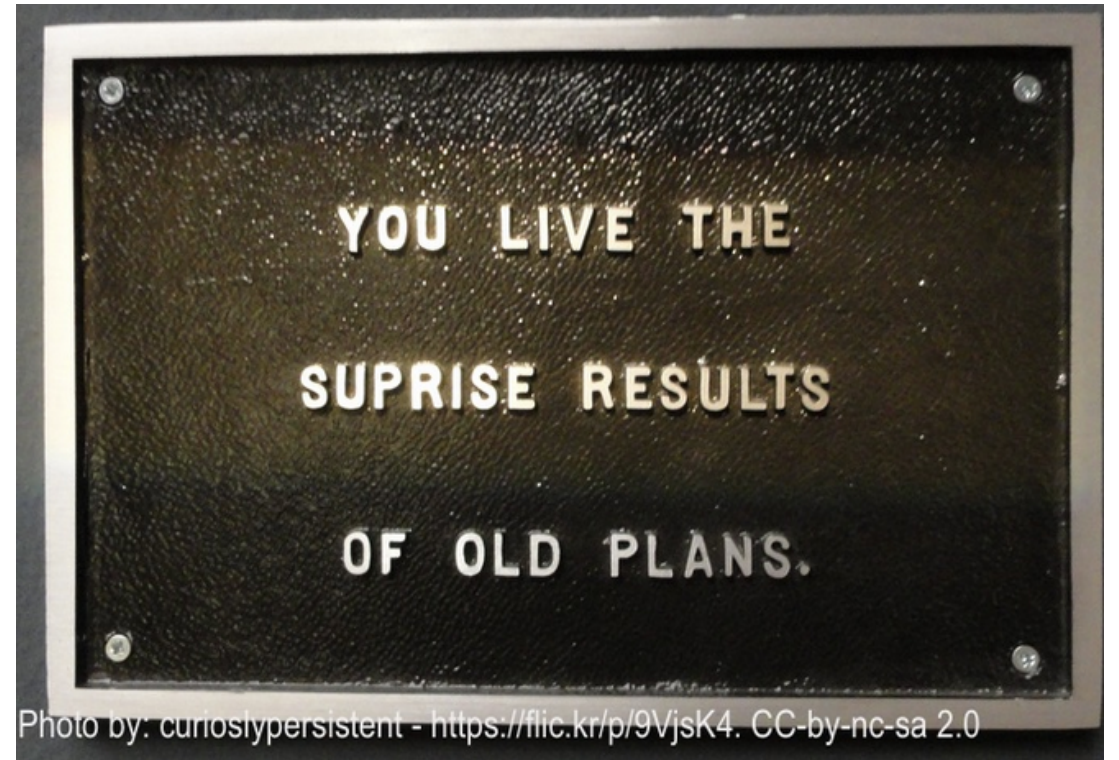 - Formats
  - Documentation

YOU LIVE THE SUPRISE RESULTS OF OLD PLANS.

*Williford, C. & Henry, C. (2012). One Culture. Computationally Intensive Research in the Humanities and Social Sciences. A Report on the Experiences of First Respondents to the Digging Into Data Challenge.

Seeing Standards. A Visualization of the Metadata Universe. Jenn Riley (2009-2010).

http://www.dlib.indiana.edu/~jenlrile/metadatamap/
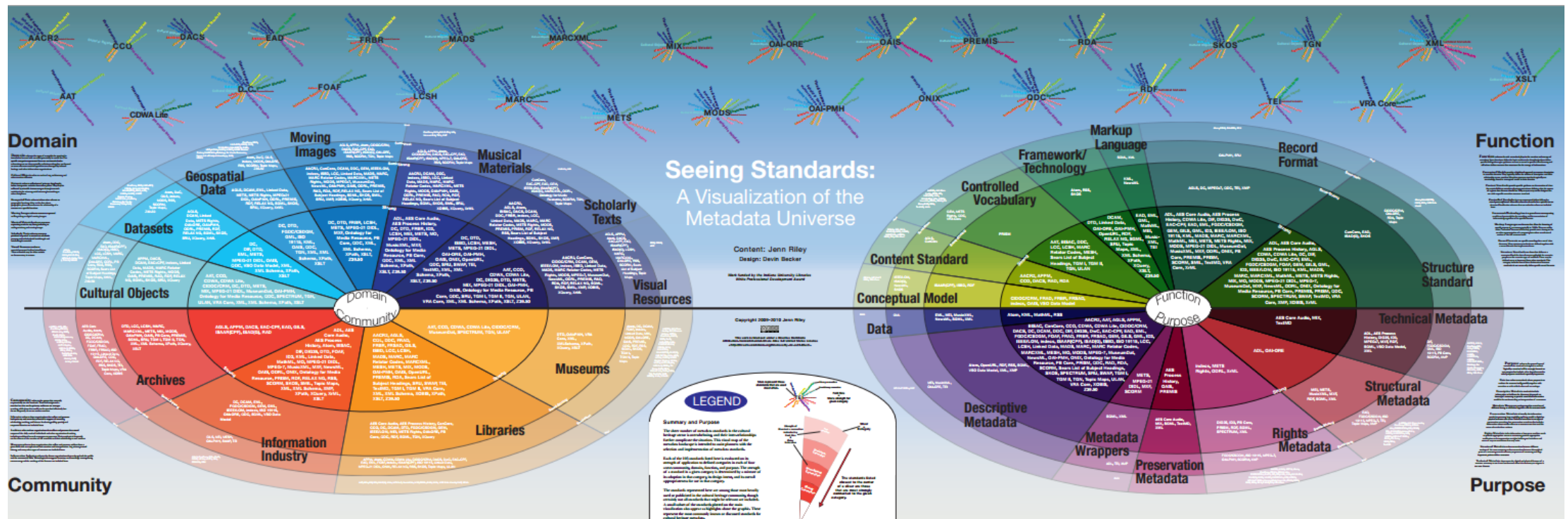
Overall Recommendations

- **Procure assistance** – consult early and often (e.g. collaborate, network)

- **Maintain documentation** from the project planning stage and throughout your work

- **Adopt a systematic model** for organizing your data: naming, file structure, formats, storage, backups

- **Adopt consistent and documented** data structures

- **Always have the entire data and research life-cycle models** in mind when you are managing your data

Data Interoperability and Linked Open Data

*The Semantic Web isn't inherently complex. The Semantic Web language, at its heart, is very, very simple. It's just about the **relationships between things**.*

Tim Berners-Lee (2007). "Q&A with Tim Berners-Lee". Bloomberg Business, April 9, 2007. http://www.bloomberg.com/bw/stories/2007-04-09/q-and-a-with-tim-berners-leebusinessweek-business-news-stock-market-and-financial-advice

*... the most important thing that was new was the idea of URI -- or URL [it was UDI back then, universal document identifier]. The idea that any piece of information anywhere should **have an identifier**, which will not only identify it, but allow you to **get hold of it**. That idea was the basic clue to the universality of the Web. That was the only thing I insisted upon.*

Tim Berners-Lee (1999). "Interview with the Web's Creator" by Chris Oakes. Wired, October 23, 1999. http://archive.wired.com/science/discoveries/news/1999/10/31830?currentPage=all

Definitions

- Interoperability
- Linked Open Data Models
- Internet Standards
    - Web Services (REST, SOAP)
- Domain Specific Standards & Protocols
    - Open Geospatial Consortium (OGC) Web Map, Web Feature and Web Converage Services (WMS, WFS, WCS)
    - DataONE, CUAHSI

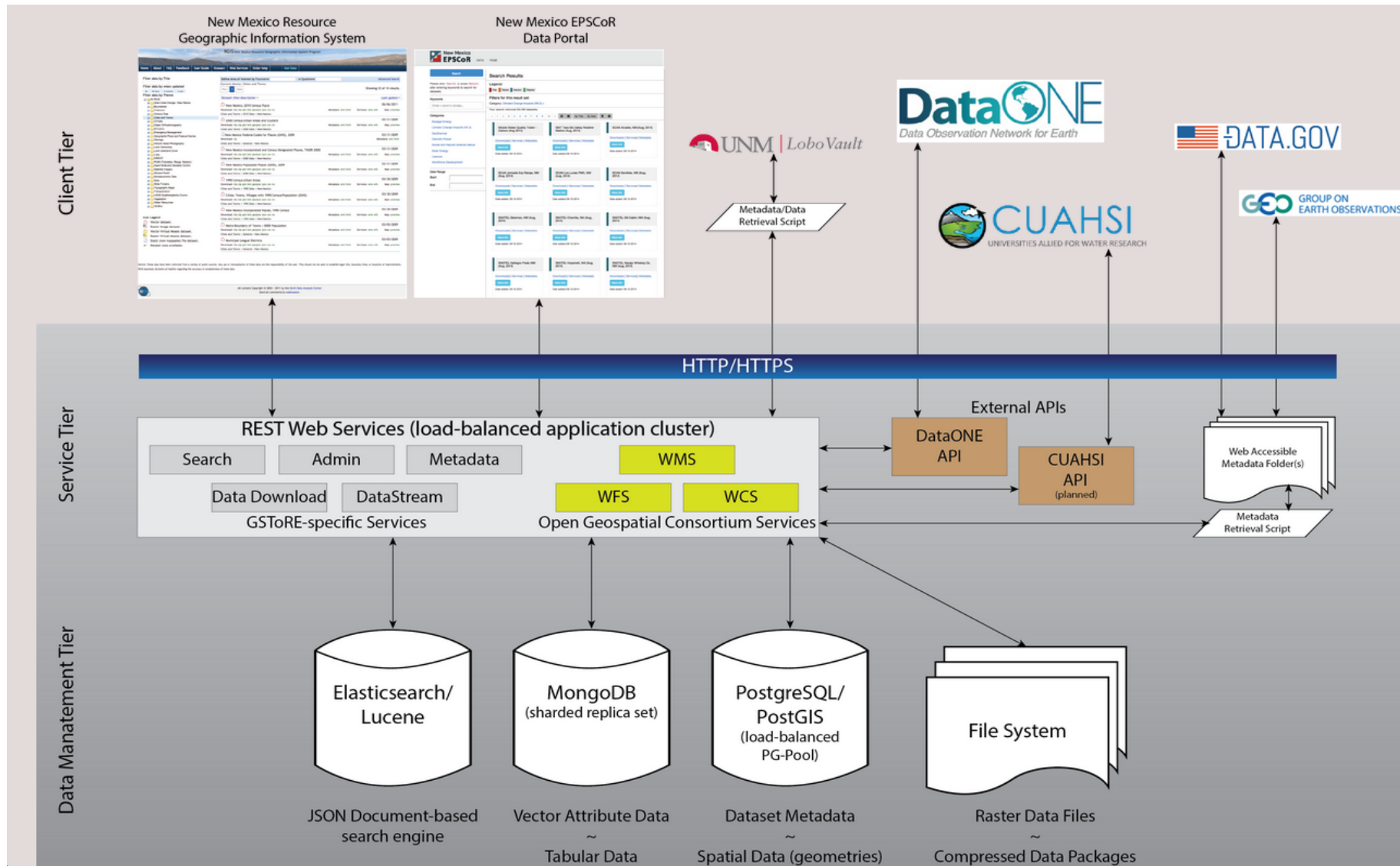★ Available on the web (whatever format) *but with an open licence, to be Open Data*

★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)

★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)

★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

★★★★★ All the above, plus: Link your data to other people's data to provide context

Tim Berners-Lee (2006). Linked Data. http://www.w3.org/DesignIssues/LinkedData.html

An Illustration

Your Data

- How does your personal experience with data management match these goals?

- What have we learned from our experiences that can inform how we communicate with and support the researchers with which we collaborate?

Photo Credit: Argonne National Laboratory. *Exponential Piles (Ref. 1-828).* https://flic.kr/p/

Acknowledgements

- CLIR Postdoctoral Fellowship Program in Data Curation at the University of New Mexico

- NSF EPSCoR Program (Track 1 [Awards: 0447691, 0814449, 1301346] and Track 2 awards [0918635, 1329470])

- New Mexico Resource Geographic Information System

- NASA ACCESS Program

- UNM's College of University Libraries and Learning Sciences