# Data & Libraries

Anna Gold, Associate Dean
Kennedy Library, Cal Poly San Luis Obispo
AGU – December 14, 2009

# Short term developments

1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)
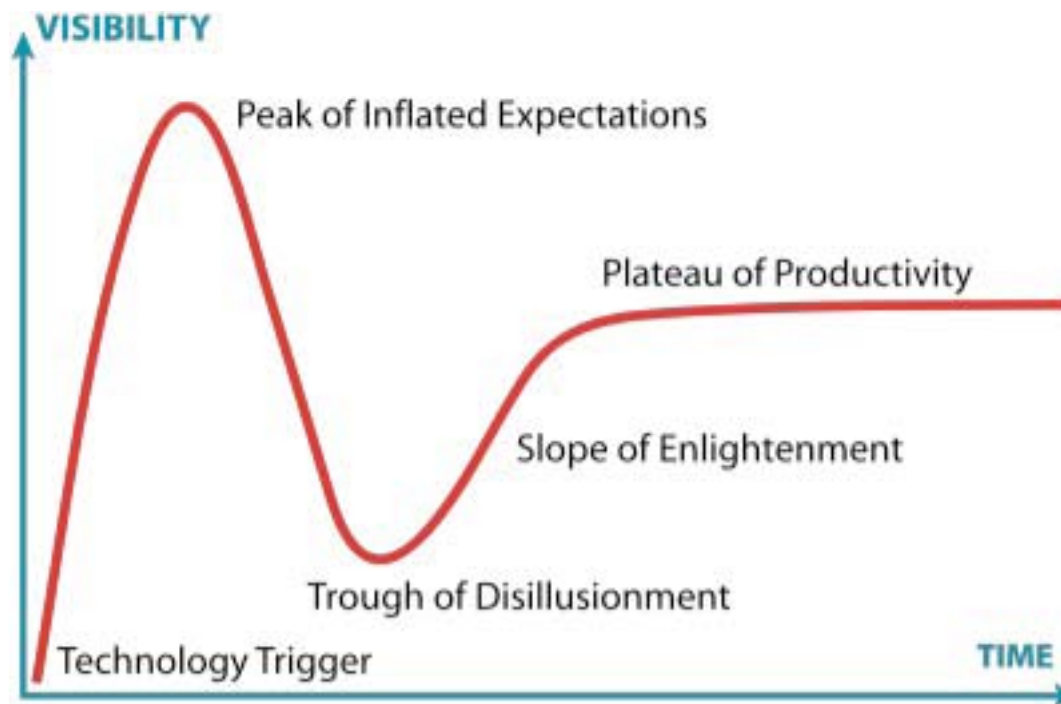
# Hype cycle?



Image from: http://en.wikipedia.org/wiki/Hype_cycle

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# Long term prospects

1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021....

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# Data curation

"Data curation is the active and ongoing management of data *through its lifecycle of interest and usefulness to scholarship, science, and education....*

# Data curation

"…which includes appraisal and selection, representation and organization of these data for access and use over time."

[Shreeves and Cragin, 2008, p. 93]

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# Data curation challenges

- Structure
- Volatility
- Scale
- Standards are lacking / inconsistent for management over time
- Global infrastructure (human, technical, economic) is lacking for storage, preservation, access over time

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# Data curation opportunities - libraries

- ☑ liaison (subject) librarians
- ☑ established repositories for research record
- ☑ scholarly communication perspective
- ☑ re-use, education perspective
- ☑ mission and values (long-term, open, neutral)
- ☑ technical knowledge (e.g. preservation, metadata)

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# Librarians are learning about data

☑ Workshops, institutes (10's)
☑ Graduate programs (3+)
☑ Certifications (1+)
☑ Conferences (10's)

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# ...and about data practices

☑ Surveys
☑ Research
☑ Interviews
☑ Case studies

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# Both top-down….

- Leadership initiatives (national, campus)
- Centers of research (Digital Curation Center)
- National policy (NRC Board)
- Graduate programs
- Certification, practitioner workshops

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# … and grassroots

- Conferences, seminars, workshops (ubiquitous)
- Communities (building, open networks)
- Pilot interviews, case studies

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# Short term developments

☑ A steady and growing record of institutional actions by library graduate schools and national library leaders to secure a long-term role for libraries in acquiring and / or stewarding collections of scientific data;

☑ Progress conceptualizing how library professionals and library-managed institutional collections of scientific research can serve the needs of science within global educational, commercial, scientific, and technological infrastructures.

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# Job posting, December 2009:

"We are seeking enthusiastic and innovative Librarian/Information Scientist to play a leadership role in <u>planning, implementing and supporting programs in data intensive research and data management</u>....initiating projects to enhance the university's research and scholarly data management and curation programs. S/he will take a lead role in developing policies and procedures for all phases of the data life cycle, data visualization support, and support of data intensive research collaboration on campus. The Librarian will also <u>develop and provide training</u>, and perform outreach to inform the campus community about the UL's data management initiatives and liaise with appropriate departments... Program development will take place in one or more of the following areas: 1) Data visualization, spatial data management, and GIS or 2) Data Intensive Research support, or 3) Data Life Cycle management...."

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# Long term prospects

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# Funding to create long-term strategies

- NSF DataNet – 5 projects, 5 yrs., $100 M
- Cost models (Blue Ribbon Task Force, in yr. 2)
- Institute for Museum & Library Services (IMLS)
- Mellon Foundation
- JISC (Joint Info. Systems Committee, UK)

# …and strategies for long-term funding?

- Blue Ribbon Task Force (US)
- Libraries?
- The existing "web infrastructure" and "SETI" model where everyone is a curator? ("Human assisted Preservation")

# Long term prospects in libraries:

☑ Legitimate role for library professionals in supporting scientific data curation.

☑ A curriculum of relevant training and education.

☑ A community of data curation professionals.

☑ Emerging consensus on means for advancing curation goals.

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# Means=Community& Ends=Science

- Means:  within and through community
- Ends: to enable the process of science, especially across disciplines and over time

# Means 1: A web of repositories

- "web of repositories" (Baker and Yarmey, 2009)
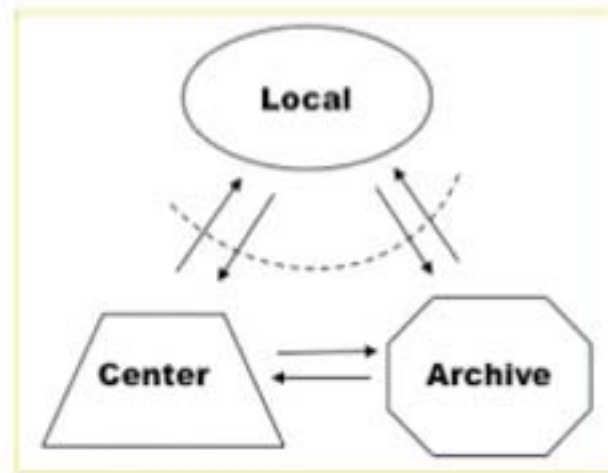


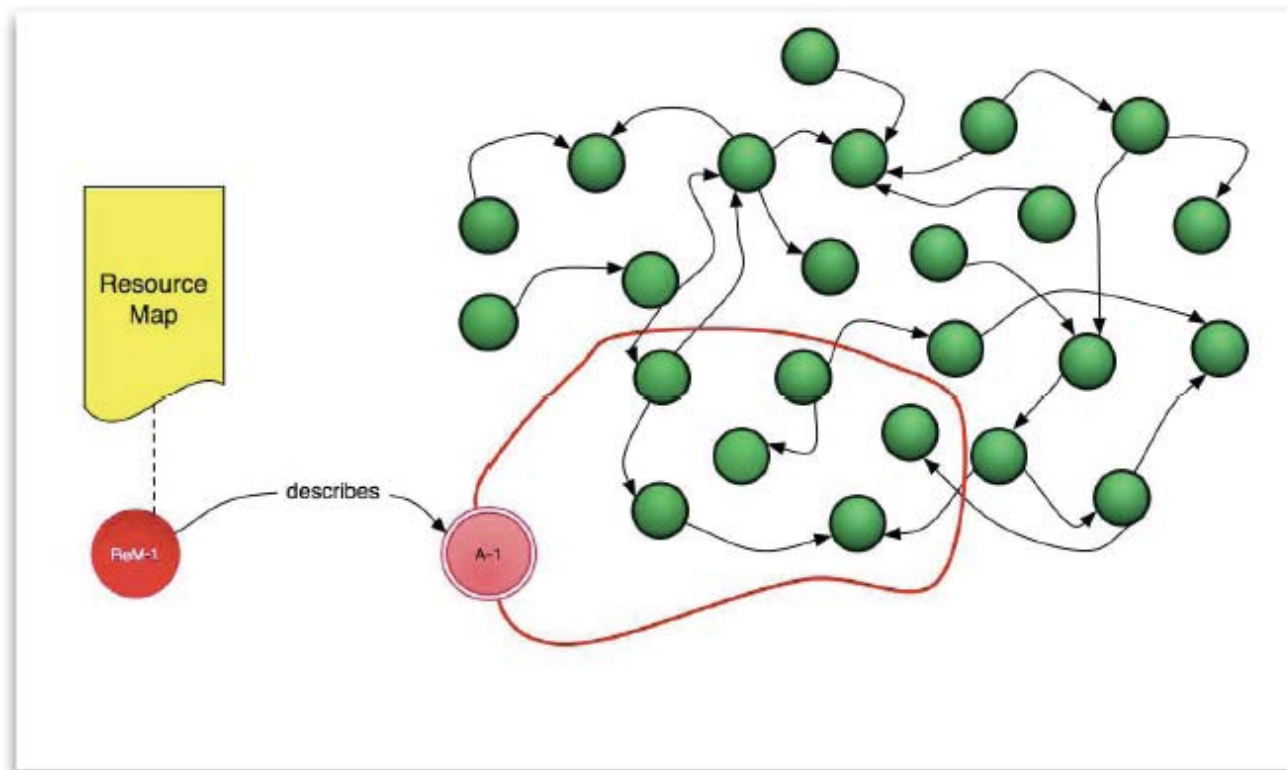- Library/institutional stewardship as a pass-through to domain-managed collections

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# Means 2: A web of digital objects

OAI-ORE – Object
Reuse and Exchange

- Bounded
  aggregations of
  digital objects (data,
  text, images, etc.)

- "Webby approach to
  compound objects" –
  simple but
  extensible, oriented
  to content and
  relationships

# Means 2:  A web of digital objects



Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)
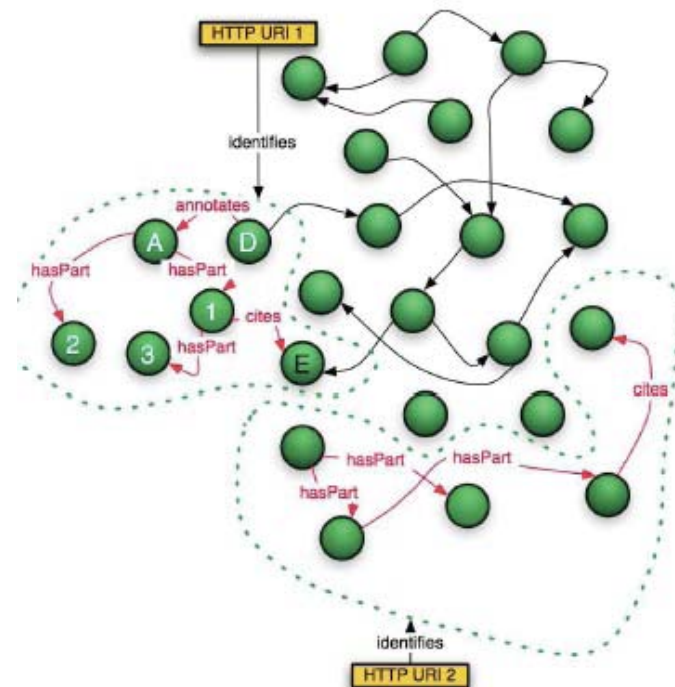
# Means 3: A web of people

⦿ Duraspace community (DSpace + Fedora)



DURASPACE™ Data Curation

Added by Thorny Staples, last edited by Thorny Staples on Oct 23, 2009 (view change)

**Data Curation Solution Community**

**Vision Statement:** Data curation is not an end, but rather a means to collect, orga
data curation will require strategic infrastructure building efforts that encompass ha
development efforts, it will be necessary to identify principles for navigation that acc
plans. Data curation should support new forms of research and learning across dis
professional and citizen researchers and learners who may also partcipate with dat

- Get Involved!
- Use Cases
- Existing Tools
- Tools That Need to be Developed
- Conferences of Interest
- Bibliography
- Organizing Group

⦿ DataNet communities ("Get involved")

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# Vision for 2015?

- Data at all scales "lives" in a highly interconnected system of social, policy, technical, and economic systems out of which basic standards and practices have emerged.
- Data research "webs" are globally supported by standards; canonical "data webs" are used in both teaching and research.
- "Dark" data sets of small science are well-managed and re-use is frequent and productive
- Data liaisons at research libraries help students and researchers participate effectively in finding, using, and managing data

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# From Geography to Topology (Emergence)
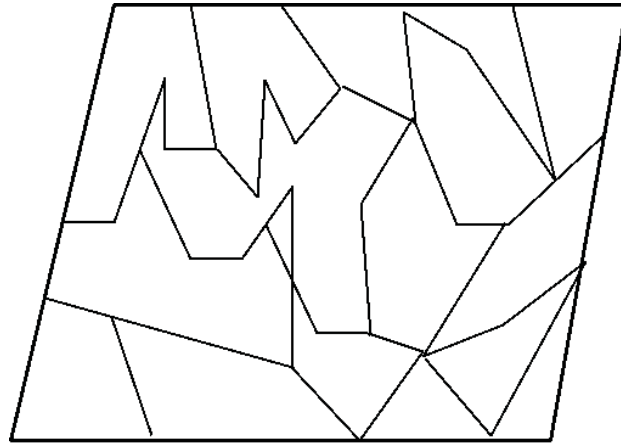


**A Logical Geography**
Map showing the structure of a set of current concepts
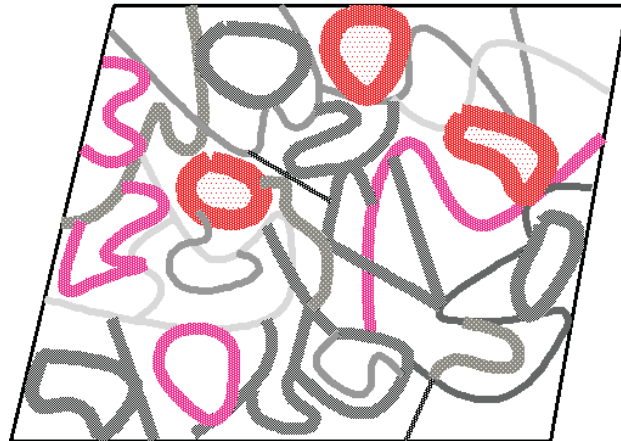
**A Logical Topography**
Map showing the structure of the corresponding part of reality as revealed by scientific and other empirical and technological advances -- subject to change as more is learnt.

One logical topography may support several possibile logical geographies.

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)

# Questions, discussion

Session on Strategies for Improved Marine and Synergistic Data Access and Interoperability (IN14A)