Title
Big Science, Little Science, and Open Science: Sustainability, Stewardship, and Knowledge Infrastructures

Permalink
https://escholarship.org/uc/item/99g200q8

Author
Borgman, Christine L.

Publication Date
2018-12-04
Big Science, Little Science, and Open Science: Sustainability, Stewardship, and Knowledge Infrastructures

Christine L. Borgman
Distinguished Research Professor
Center for Knowledge Infrastructures
University of California, Los Angeles
http://christineborgman.info
https://knowledgeinfrastructures.gseis.ucla.edu
@scitechprof

Keynote Presentation
National Open Science Plan for France: From Strategy to Action
Paris, 4 December 2018
• 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
“robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds”

Knowledge Infrastructures
Infrastructures

Infrastructures

Opportunities in Open Science
New knowledge from old data

- https://library.cfa.harvard.edu/image-vocab/harvard-computers
Networks of data

http://humannaturelab.net/wp-content/uploads/2015/01/Fig1-no-text-village-2-only-selection.png
The Undiscovered: Many great discoveries in science are surprises.

https://www.radcliffe.harvard.edu/event/2018-undiscovered-symposium
Challenges in Open Science
Data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship.


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)
Science <-> Data

Engineering researcher: "Temperature is temperature."

CENS Robotics team
Science <-> Data

Engineering researcher: “Temperature is temperature.”

Biologist: “There are hundreds of ways to measure temperature. ‘The temperature is 98’ is low-value compared to, ‘the temperature of the surface, measured by the infrared thermopile, model number XYZ, is 98.’ That means it is measuring a proxy for a temperature, rather than being in contact with a probe, and it is measuring from a distance. The accuracy is plus or minus .05 of a degree. I [also] want to know that it was taken outside versus inside a controlled environment, how long it had been in place, and the last time it was calibrated, which might tell me whether it has drifted.."
Project Timelines

LSST Timeline: https://www.lsst.org/about/timeline
Data Stewardship

Marie Curie’s notebook aip.org

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

Marie Curie’s notebook aip.org

User

Application Software

Operating System

Hardware

Wikipedia.org

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

Pisa Griffin

Date:1/2/07 75  Place: Sakaltutan Zafir
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. So 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 Zafir
Household now Zafir and wife; Nazif Unal and wifine 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. Zafir then said, keskin değild (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.
The DCC Curation Lifecycle Model

**The DCC Curation Lifecycle Model**

**1. CONCEPTUALISE**
- **Data (Digital Objects or Databases)**
- **Description**
- **Representation Information**
- **Preservation Planning**

**2. CREATE OR RECEIVE**
- **Curate**
- **Preserve**
- **Preservation Action**
- **Ingest**
- **Store**

**3. APPRAISE & SELECT**
- **Access, Use & Reuse**

**4. TRANSFORM**

**5. DISPOSE**
- **Dispose**
- **Reappraise**
- **Migrate**

---

**Data, any information in binary digital form, is at the centre of the Curation Lifecycle. This includes:**

- **Simple Digital Objects** are discrete digital items; such as textual files, images or sound files, along with their related identifiers and metadata.
- **Complex Digital Objects** are discrete objects made by combining a number of other digital objects, such as websites, structured collections of records or data stored in a computer system.

---

**Full Lifecycle Actions**

**Sequential Actions**

1. CONCEPTUALISE
2. CREATE OR RECEIVE
3. APPRAISE & SELECT
4. TRANSFORM
5. PRESERVE
6. INGEST
7. STORE
8. CURATE
9. ACCESS, USE & REUSE
10. DISPOSE

**Occasional Actions**

1. Dispose of data, which has not been selected for long-term curation and preservation in accordance with documented policies, guidance or legal requirements. Typically data may be transferred to another archive, repository, data centre or other custodian. In some instances data is destroyed. The data's nature may, for legal reasons, necessitate secure destruction.
2. Return data which fails validation procedures for further appraisal and reselection.
3. Migrate data to a different format. This may be done to accord with the storage environment or to ensure the data's immunity from hardware or software obsolescence.

---

The DCC Curation Lifecycle Model provides a graphical high level overview of the stages required for successful curation and preservation of data from initial conceptualisation or receipt. The model can be used to plan activities within an organisation or consortium to ensure that all necessary stages are undertaken, each in the correct sequence. The model enables granular functionality to be mapped against it; to define roles and responsibilities, and build a framework of standards and technologies to implement. It can help with the process of identifying additional steps which may be required, or actions which are not required by certain situations or disciplines, and ensuring that processes and policies are adequately documented.

---

**www.dcc.ac.uk**

**info@dcc.ac.uk**
Data Stewardship: the Reality

We just need to migrate the data from these systems to fit into that hole over there.

I’ll get the hammer.

Graduate students

Post-doctoral fellows

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


http://gsa.rice.edu/

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

Mount Wilson Solar Observatory, 2017

Gettys Research Institute


Graduate students

Post-doctoral fellows

http://www.gsa.rice.edu/

https://med.nyu.edu/our-community/life-nyu-school-medicine/life-postdoc
Data Stewardship: The Ideal

Pasquetto, I.V. (2018). *From Open Data to Knowledge Production: Biomedical Data Sharing and Unpredictable Data Reuses*. PhD Dissertation. [https://escholarship.org/uc/item/1sx7v77r](https://escholarship.org/uc/item/1sx7v77r)
'This is just the beginning': Using DNA and genealogy to crack years-old cold cases

Police are harnessing consumer DNA sites to solve old murders, which could spur a massive clearing of unsolved crimes.

by Kate Snow and Jon Schuppe / Jul 18, 2018 / 4:30 AM ET

Genealogy databases and the future of criminal investigation

Natalie Ram¹, Christi J. Guerrini², Amy L. McGuire²
+ See all authors and affiliations

DOI: 10.1126/science.aau1083

Paths Toward Open Science
Opportunities and Challenges

• Opportunities
  – Capture scientific products in digital form
  – Store, integrate, generate new knowledge

• Challenges
  – Skills and resources required to curate scientific records
  – Career paths for data science, curation work
  – Sustainability and stewardship of scientific products
  – Uses, reuses, and misuses of scientific products
Sustainable Open Science

• Create career paths
  – Data science
  – Curation and stewardship

• Commit to long-term infrastructure investments
  – Capture and sustain scholarly products
  – Stewardship of knowledge infrastructures

• Promote data reuse
  – Celebrate discovery
  – Anticipate controversy
  – Govern misuse