Data Curation with the UCSD Library

Rick Wagner

May 22, 2012

Synopsis

- UCSD is putting effort into shared IT
- Data curation services provided to support funding agency mandates
  - (e.g., NSF Data Management)
- 5 pilot projects were chosen to work through the process
- Our simulations were chosen because of my work in the IVOA
- I chose to participate to learn about how the library thinks about data and metadata

UCSanDiego
Local Impact, National Influence, Global Reach

Research Cyberinfrastructure (RCI)

"Making connections is fundamental to scientific research.

RCI offers UC San Diego researchers the computing, network, and human infrastructure needed to create, manage, and share data. Principal investigators are encouraged to use the campus's RCI in addressing federal sponsors' existing and new data management requirements. SDSC's favorable pricing can help researchers justify costs associated with data generation and management—costs that are increasingly included in proposal budgets."

UCSanDiego
Local Impact, National Influence, Global Reach

A brief history …

- 2008: Campus-wide needs assessment

- April, 2009: *Blueprint for the Digital University* issued

- April, 2010: *Cyberinfrastructure Planning and Operations Committee Report* issued
  - report accepted, business plan created

- January, 2011: RCI Oversight Committee formed
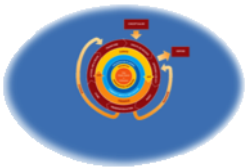 - business plan accepted, oversight committee charged

UCSanDiego
Local Impact, National Influence, Global Reach

RCI elements

- High-Performance Computing

- Data Center Colocation

- Storage

- Networking and other services

- Data curation

UCSD research cyberinfrastructure

UCSanDiego
Local Impact, National Influence, Global Reach

Data curation

- Starting with a two year pilot phase

- Using existing tools whenever possible
  - Storage at SDSC
  - Digital Asset Management System at UCSD Libraries
  - Campus high-speed networking
  - Chronopolis digital preservation network
  - DOIs from EZID
  - Metadata published to the Online Archive of California
    (http://www.oac.cdlib.org/)

http://rci.ucsd.edu/services/data-curation.html

http://rci.ucsd.edu/pilots

UCSanDiego
Local Impact, National Influence, Global Reach

The Brain Observatory

Preserve and curate the digital version of the brain of patient HM, the most studied neuropsychological patient in modern medicine.
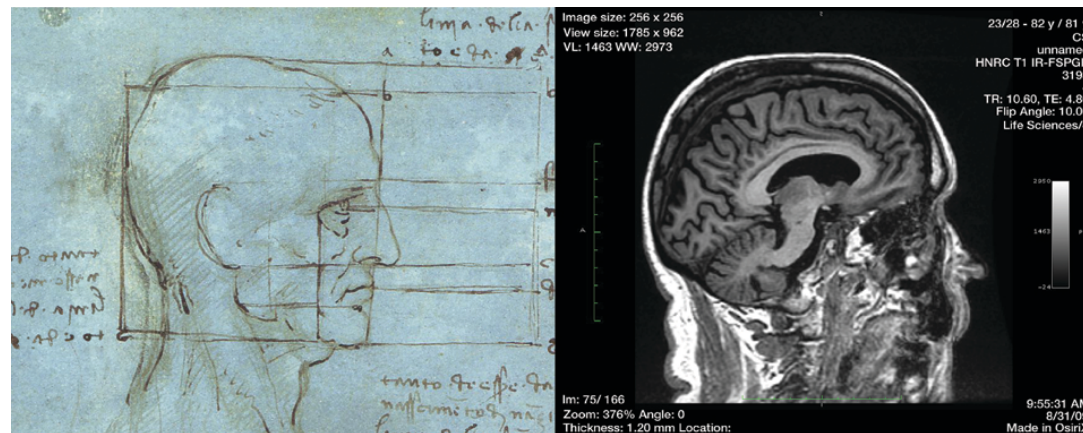
NSF OpenTopography Facility

Provide fully curated, DOI-referenced, long-term hosting for the OpenTopography data archive

Levantine Archaeology Laboratory

Develop the infrastructure needed to curate cultural heritage data that is spurred by new visualization and analysis tools.

Scripps Institution of Oceanography Geological Collections

Assist with the creation of a standards-based access, discovery and preservation system for one of the largest collections of marine geology samples in the United States.
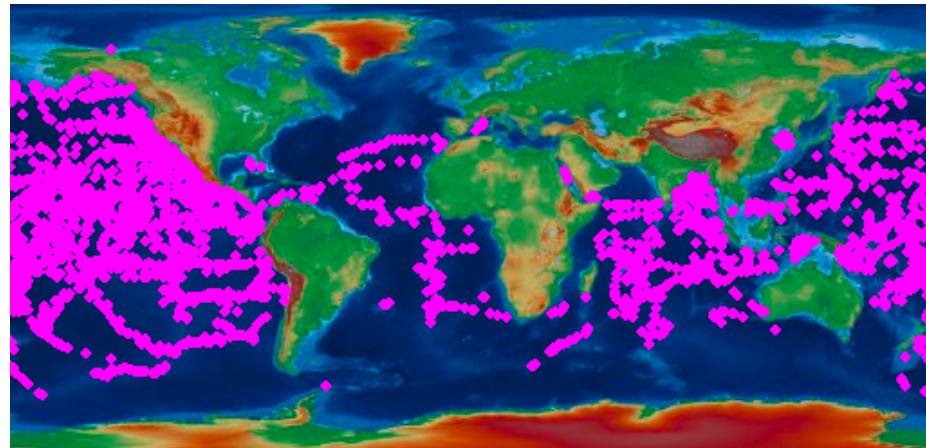
The Laboratory for Computational Astrophysics

Provide data management and curation to improve collaborations with other UCSD researchers, and to support publishing simulations of astrophysical phenomenon in cosmology, star formation and turbulence.

UCSanDiego
Local Impact, National Influence, Global Reach

Now for some opinions

- Library roles are changing

- Campus libraries are natural partners in curation

- Library metadata is good at describing domains
  - i.e., this is biology, it has plants, puppies, and DNA

- But a lot gets lost in translation
  - Like that Simulation Data Model I worked on

- The files will be available via HTTP, the other interfaces are unclear

- The IVOA is a unique organization
  - Other domains do not speak in terms of registries and standardized services and publishing

Yesterday's Email

SUBJECT: Update on Library Consolidations

Like many academic libraries, the UC San Diego Libraries are witnessing increased demand for digital services and resources and declining demand for physical resources.

...

Brian E. C. Schottlaender
The Audrey Geisel University Librarian

UCSanDiego
Local Impact, National Influence, Global Reach