



DC-2009
SEOUL

International Conference on Dublin Core and Metadata Applications

12-16 October 2009, SEOUL, KOREA



DC Science and Metadata Community Meeting

Jane Greenburg

University of North Carolina
School of Information and Library Science

Stuart Weibel

OCLC Research



DC - SAM

- Chartered to bring together global metadata expertise in support of the management of scientific data.
- In particular:
 - Policies and technology to support persistent canonical identification of data
 - Metadata to support discovery and management of data sets so as to make them more discoverable and to enhance reuse



Global awareness of digital curation issues is rising rapidly

- DCC (Digital Curation Center) – UK
 - <http://digitalcuration.blogspot.com/>
- NSF Blue Ribbon Panel on Digital Preservation
 - <http://brtf.sdsc.edu/>
- NCDD – Netherlands Coalition for Digital Preservation
 - <http://www.ncdd.nl/en/index.php>
- PADI & ICADS:
 - <http://www.nla.gov.au/padi/index.html>



Scientific Data Curation has special problems of its own

- Scientific Data Sets
 - Non-textual
 - Many formats
 - Not self-declarative
 - Quantity has a Quality all its own
 - High semantic barriers across domains
 - Often not designed for reuse and repurposing



Harnessing the Power of Digital Data for Science and Society

- Central Whitepaper on digital data issues
 - Key Characteristics of the current Digital Data Landscape
 - Guiding Principles
 - Vision
 - Strategy
- http://www.nitrd.gov/about/Harnessing_Power_Web.pdf



Key characteristics of the current digital data landscape

- the products of science and the starting point for new research are increasingly “born-digital”
- exploding volumes and demand for data are driven by the rapid pace of digital innovations
- all sectors of society are stakeholders in digital preservation and access;
- a comprehensive framework for managing digital data is missing.



Guiding Principles

- Science is global, and digital scientific data are national and global assets
- Not all digital scientific data need to be preserved, or preserved indefinitely (selection!)
- Communities of practice are an essential feature of the digital landscape
- Preservation of digital scientific data is both a government and private sector responsibility
- Preservation, access, and interoperability require management of the full data life cycle
- Dynamic strategies are required



Strategy

- *Create a **comprehensive framework** of transparent, evolvable, extensible **policies** and management and **organizational structures** that provide **reliable, effective access** to the full spectrum of public digital scientific data.*



The NSF DataNet solicitation is the tip of the spear in the US

- 5 year, \$100,000,000 initiative
- 5 projects (2 in Round 1, 3 in Round 2)
- Promote new organizations and innovative technology to support sustainable data curation



Cliff Lynch

- Data stewardship is a clean slate, can be engineered from scratch
- Bit-level management is distributed, with business models driving costs towards commoditization
- Who will add organizational and management value? Libraries? IT departments? Consortia? Change is afoot: opportunity and threat for libraries
- Data as services rather than formats
- High-level curation is currently patchwork. How should it be funded, organized, managed?
- Sustainability is a hard problem



To Join DC-SAM

- Visit the DC-SAM community page at:
<http://dublincore.org/groups/sam/index.shtml>
- The DC-SAM Wiki is at:
http://ils.unc.edu/spaces/sam/index.php/Main_Page