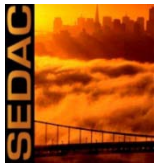
THE EARTH INSTITUTE
COLUMBIA UNIVERSITY



Sustainable Governance for Long-Term Stewardship of Earth Science Data

Robert R. Downs and Robert S. Chen

NASA Socioeconomic Data and Applications Center (SEDAC)
Center for International Earth Science Information Network (CIESIN)
The Earth Institute, Columbia University

Prepared for presentation to the
Earth and Space Science Informatics Workshop
August 4, 2010
George Mason University
Fairfax, Virginia

Abstract

In addition to technical infrastructure and capabilities, the long-term management of Earth science data requires **organizational sustainability** to provide continuing stewardship to address the risks to scientific data and support their use by future communities. Providing sustainable infrastructure for the preservation of scientific data requires **organizational commitments, capacity, structures, and plans for data stewardship** that are consistent with the missions of the organizations that accept the responsibility to serve in data stewardship roles. Alternative approaches to attaining organizational sustainability for interdisciplinary human dimensions and polar data are discussed in terms of recent recommendations for organizational sustainability to foster digital preservation.

Where Do Old Data Go?

- Scientific data are often collected through projects or missions with limited lifetimes
  - Funding received to collect data for a specific period
  - Limited lifetime of instrument or platform or limited time period for data collection and processing
- Data may be at risk after project or mission ends
  - Funds for project activities expire and project team disperses
  - Resources often not available to clean up and properly document and archive project data
  - Those most knowledgeable about the data become inaccessible, e.g., due to changes in responsibilities, jobs, retirement, illness, or death
  - Computer and information systems are replaced and not all contents are migrated
  - Active archives lack sufficient resources to track, access, and archive important datasets

Organizational Needs for Sustainable Stewardship of Scientific Data

- Organizational commitment to long-term scientific data stewardship and associated costs
- Capacity to manage scientific data on an ongoing basis
- Organizational structure for allocating resources and staff to plan and perform scientific data stewardship
- Plans for data stewardship that are consistent with the mission of the organization

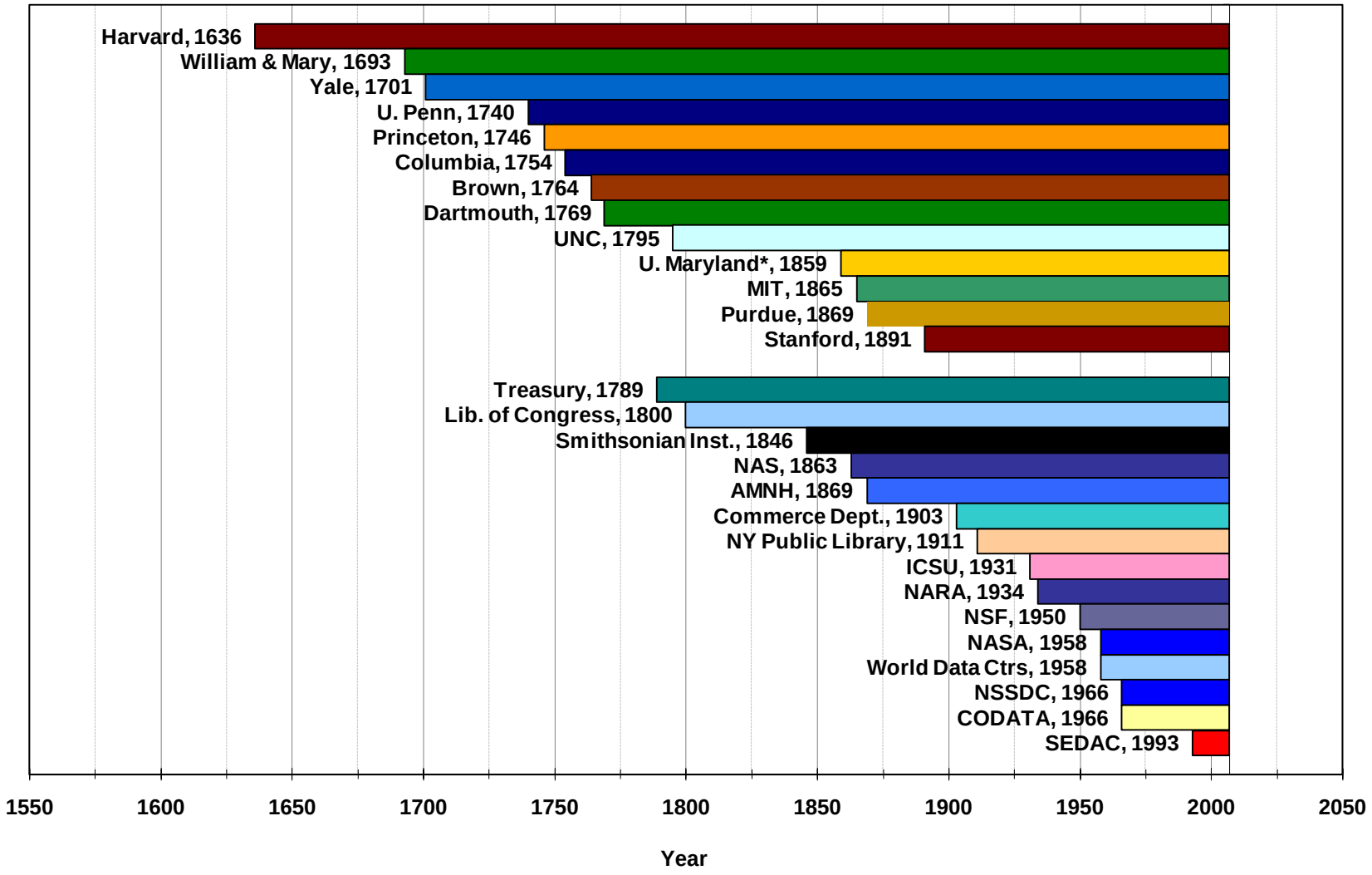
Traditional Organizational Models for Sustainable Data Stewardship

- Government repositories
  - Government agency or institution mission includes data stewardship
  - Budget allocated based on mission needs; sometimes includes cost recovery, user fees
  - Data management capacity deployed and operated by government or contractors
  - Examples: Library of Congress, National Archives and Records Administration, Smithsonian Institution, USGS, NOAA, US Census Bureau
- Non-government repositories and networks
  - Data managed by public or private nongovernmental institutions, including universities
  - Funding from a mix of sources, including depositor, membership, and user fees; grants and contracts; gifts and endowments; volunteer and educational activities; publication and other value-added services
  - Examples: University depositories, museum collections, ICPSR, IRIS

Concerns about Traditional Organizational Models

- Government repositories
  - Missions and budgets subject to change based on legislative or executive decision making, which may put short-term issues ahead of long-term needs
  - Longevity of agencies not guaranteed…only a few Federal agencies with more than a century of successful experience in records management and retention
  - Holdings could be subject to loss or corruption due to political influence
- Non-government repositories and networks
  - Missions and budgets not necessarily sustainable due to changes in economic, social or political conditions and user needs and expectations
  - Traditional funding models at risk, e.g., support for physical infrastructure vs. digital infrastructure, open access vs. subscription models, ability to subsidize preservation through educational or other value-added activities

Examples of Organizational Longevity



Based on Downs and Chen (2010), *Journal of Digital Information*
http://journals.tdl.org/jodi/article/view/753

Potential Alternatives to Traditional Models

- "Hybrid" government-private model
  - Establishment of public-private institutions or consortia with combination of government support and private infrastructure and independence
  - Development of business models that generate resources from combinations of sources, e.g., endowments, gifts, grants, subscriptions, fees, royalties, value-added support, micropayments (iTunes), etc.

- "Information Commons" approach
  - Development of communities that share scientific data and agree to serve as data stewards in an organized, voluntary framework
  - Business model based on combination of voluntary efforts, distributed organizational and institutional support, and other gifts, endowments, etc.
  - Not yet clear how an information commons can meet the needs for long-term organizational commitment, data management capacity, and plans

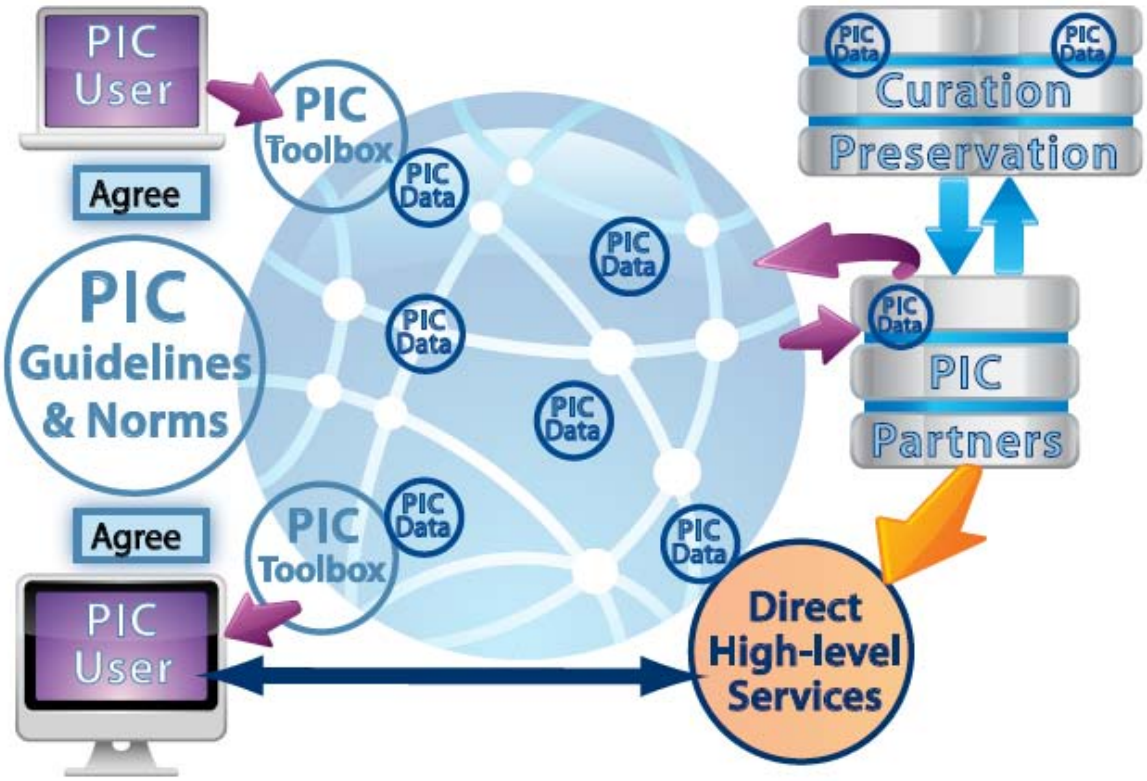
Polar Information Commons (PIC)

- International initiative launched in June 2010 to provide long-term stewardship of polar science data
- Core approach is based on the Science Commons Protocol for Implementing Open Access Data:
  - http://www.sciencecommons.org/projects/publishing/open-access-data-protocol/
- Data owners/originators use the Creative Commons Zero (CC0) license to place their data in the public domain or use CC-BY-3.0 where copyrights apply
- The PIC has developed a digital label (the “PIC badge”) and a set of community norms to facilitate but not require data use and reuse

http://www.polarcommons.org

PIC Stewardship Approach

- Polar data sources expose their data to the world through the PIC badge and open protocols
- Data centers monitor new PIC data and assess and acquire important data for formal archiving, curation, and access
- The PIC community encourages compliance with PIC norms through peer pressure, brand identity, arrangements with journals and funding



- The PIC community provides useful tools for PIC data quality, search, access, documentation, attribution, metrics, etc.

Key Challenges for the PIC

- Can the PIC develop sustainable collaboration between:
  - Scientists and scientific bodies to support the “commons” (content, tools, and rights)?
  - Universities and other digital data repositories to support selection and appraisal and long-term institutional commitments to knowledge preservation and dissemination?
  - Governments and other governmental bodies to provide resources and requirements (carrots and sticks)?
  - The broader public to provide volunteer labor, infrastructure, and other support?
- How can a reliable virtual organizational framework and infrastructure for the PIC be developed and maintained over the long term?

Summary

- Long-term data stewardship requires a long-term commitment to data management and preservation
- Data stewardship must be consistent with long-term organizational objectives
- A long-term but flexible business model is needed to ensure that resources are available in the long term to support data stewardship activities
- An organization's history is important for understanding its ability to adapt and survive – but is not a blueprint for its future!
- New organizational approaches and models are needed in the new digital age – the status quo won't last much longer!