

Version 7.0 August 10, 2011

**INTRODUCTION**

This is a compilation of data lifecycle models and concepts assembled in part to fulfill

Committee on Earth Observation Satellites (CEOS) Working Group on Information Systems and Services (WGISS) and the U.S. Geological Survey (USGS) Community for Data Integration Data Management Best Practices needs. It is intended to be a living document, which will evolve as new information is discovered.

**CONTENTS**

1. Digital Curation Centre (DCC) Lifecycle Model
2. Ellyn Montgomery, USGS, Data Lifecycle Diagram
3. FGDC Stages of the Geospatial Data Lifecycle pursuant to OMB Circular A–16
4. University of Oxford Research Data Management Chart
5. NOAA Environmental Data Life Cycle Functions
6. Open Archival Information System (OAIS) Framework
7. USGS Scientific Information Management Workshop Vocabulary
8. Peter Fox Lifecycle Diagrams
9. National Science Foundation
10. NDIIPP Preserving Our Digital Heritage
11. What Researchers Want
12. EPA Project Life Cycle
13. IWGDD’s Digital Data Life Cycle Model
14. Scientific Data Management Plan Guidance
15. Linear Data Life Cycle
16. Generic Science Data Lifecycle
17. Cassandra Ladino Hybrid Data Lifecycle Model
18. Ray Obuch Data Management – A Lifecycle Approach
19. USGS Data Management Plan Framework (DMPf) – Smith, Tessler, and McHale
20. BLM Data Management Handbook
21. ARL Joint Task Force on Library Support for E-Science
22. U.S. Department of Health and Human Services Key Components
23. ICPSR Preservation over the Data Life Cycle
24. William Michener DataONE: Data Life Cycle Management
25. IBM Aspects of Lifecycle Management - Research
26. University of California San Diego Digital Curation Program
27. University of Miami Scientific Data Lifecycle
28. Managing Research Data Lifecycles through Context
29. Purdue University Research Repository Collaborative Model
30. Data Lifecycle Management at Dell
31. John Faundeen & Ellyn Montgomery “Spins”
32. DataTrain Model 1
33. DataTrain Model 2
34. Steve Tessler Data Management Cycles
35. Peter Fox Full Life Cycle of Data
36. Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle
37. Steve Tessler Data and System Lifecycle Models
38. University of Virginia Library
39. Mario Valle
40. Michigan State University Records Life Cycle Model
41. GeoMAPP Geoarchiving Process Lifecycle
42. Jeff de La Beaujardière's proposed Data Management Framework
43. Sarah Demb Continuumm Model
44. USGS Data Life Cycle Model Elements

**1.** **THE DIGITAL CURATION CENTRE MODEL**

The Digital Curation Centre is based in Great Britain. The URL for the information

below is http://www.dcc.ac.uk/resources/curation-lifecycle-model

Our 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. You can use our model to plan activities within your organisation or consortium

to ensure that all of the necessary steps in the curation lifecycle are covered. It is

important to note that the model is an ideal. In reality, users of the model may enter at

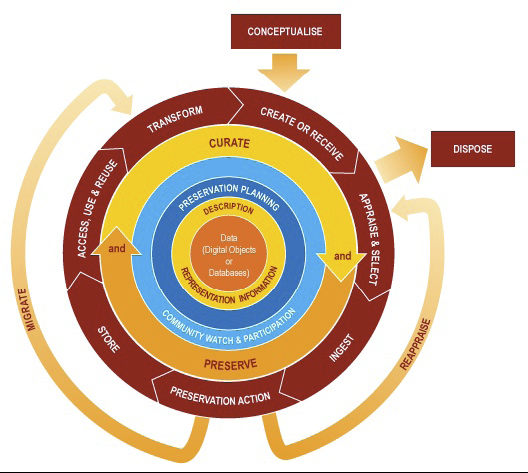
any stage of the lifecycle depending on their current area of need. For instance, a digital

repository manager may engage with the model for this first time when considering

curation from the point of ingest. The repository manger may then work backwards

to refine the support they offer during the conceptualisation and creation processes to

improve data management and longer-term curation.



**Key elements of the DCC Curation Lifecycle Model Data**

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

**Digital Objects:** simple digital objects (discrete digital items such as text files, image files or sound files,

along with their related identifiers and metadata) or complex digital objects (discrete digital objects made

by combining a number of other digital objects, such as websites)

**Databases:** structured collections of records or data stored in a computer system

**Full Lifecycle Actions**

**Description and Representation Information**

Assign administrative, descriptive, technical, structural and preservation metadata, using appropriate

standards, to ensure adequate description and control over the long-term. Collect and assign representation

information required to understand and render both the digital material and the associated metadata.

**Preservation Planning**

Plan for preservation throughout the curation lifecycle of digital material. This would include plans for

management and administration of all curation lifecycle actions.

**Community Watch and Participation**

Maintain a watch on appropriate community activities, and participate in the development of shared

standards, tools and suitable software.

**Curate and Preserve**

Be aware of, and undertake management and administrative actions planned to promote curation and

preservation throughout the curation lifecycle.

**Sequential Actions**

**Conceptualise**

Conceive and plan the creation of data, including capture method and storage options.

**Create or Receive**

Create data including administrative, descriptive, structural and technical metadata. Preservation metadata

may also be added at the time of creation.

Receive data, in accordance with documented collecting policies, from data creators, other archives,

repositories or data centres, and if required assign appropriate metadata.

**Appraise and Select**

Evaluate data and select for long-term curation and preservation. Adhere to documented guidance, policies

or legal requirements.

**Ingest**

Transfer data to an archive, repository, data centre or other custodian. Adhere to documented guidance,

policies or legal requirements.

**Preservation Action**

Undertake actions to ensure long-term preservation and retention of the authoritative nature of data.

Preservation actions should ensure that data remains authentic, reliable and usable while maintaining its

integrity. Actions include data cleaning, validation, assigning preservation metadata, assigning

representation information and ensuring acceptable data structures or file formats.

**Store**

Store the data in a secure manner adhering to relevant standards.

**Access, Use and Reuse**

Ensure that data is accessible to both designated users and reusers, on a day-to-day basis. This may be in

the form of publicly available published information. Robust access controls and authentication procedures

may be applicable.

**Transform**

Create new data from the original, for example:

by migration into a different format, or

by creating a subset, by selection or query, to create newly derived results, perhaps for publication

**Occasional Actions**

**Dispose**

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.

**Reappraise**

Return data which fails validation procedures for further appraisal and re-selection.

**Migrate**

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.

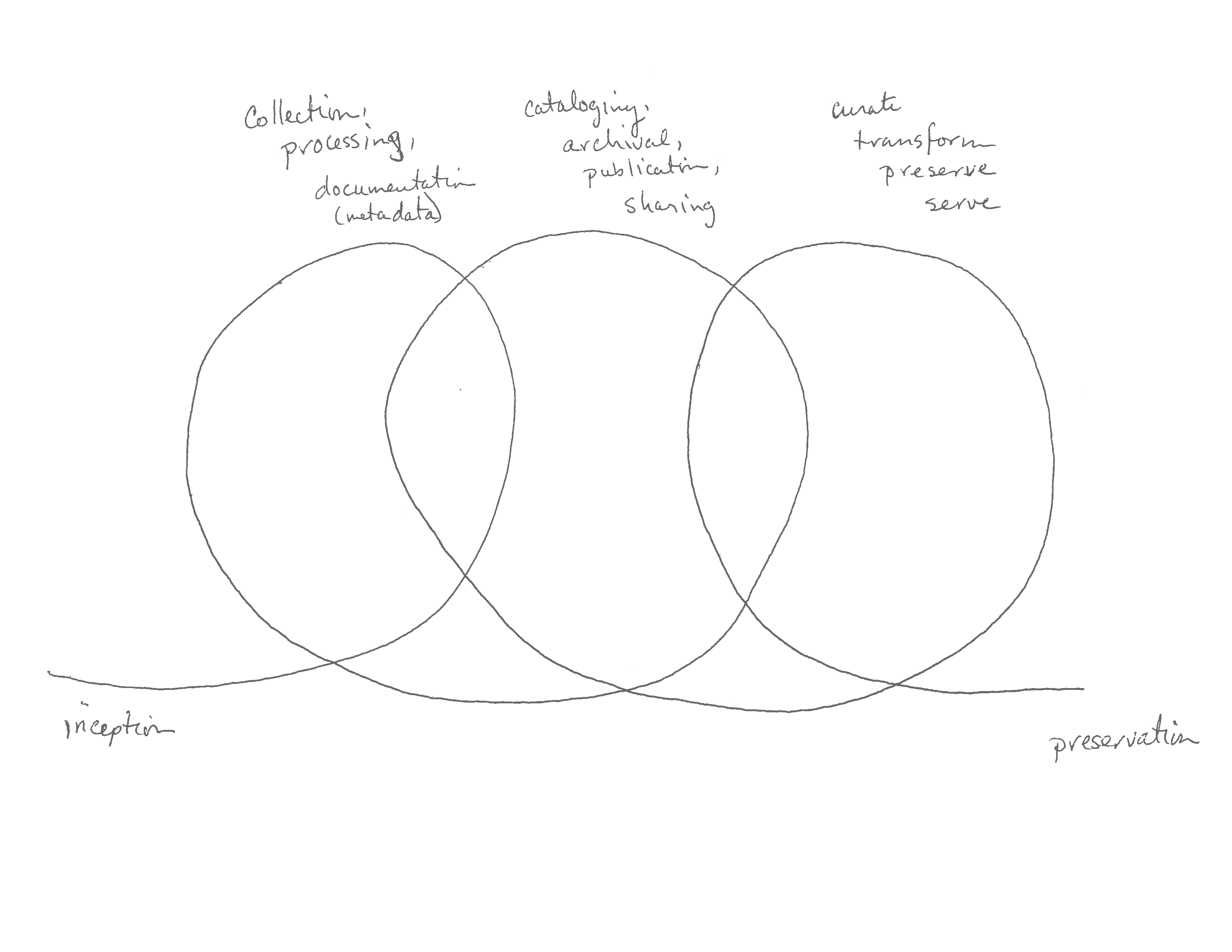
**2. THE ELLYN MONTGOMERY, USGS, DATA LIFECYCLE DIAGRAM**

Provided via email 18 November 2010.

Many of the diagrams out there have appealing elements, but I couldn't find one I really liked so sketched

something that's a combination of some of Ted Habermann's ideas and the material at:

http://imageweb.zoo.ox.ac.uk/wiki/index.php/ADMIRAL\_Data\_Management\_Plan\_Template.



The ideas I tried to capture are:

1) it's a non-linear (and perhaps multi-threaded) process

2) multiple loops or phases (not restricted to the number drawn) that may overlap are needed

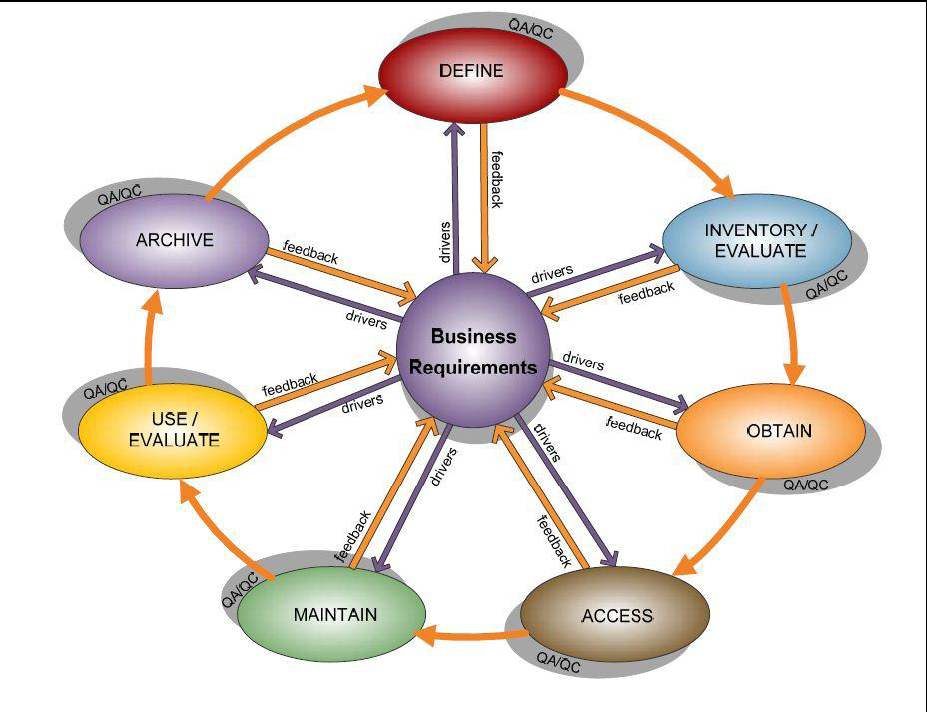
3) parts of the process are ongoing

4) there's a transition between data provider and data curator somewhere in the middle of the

progression this may vary between types of data and the eventual avenue for publication and distribution

**3. FGDC Stages of the Geospatial Data Lifecycle pursuant to OMB Circular A–16**

Figure A1 below shows the FGDC lifecycle model, which advocates compliance of Office of Management and Budget (OMB) Circular A-16, “Coordination of Geographic Information and Related Spatial Data Activities.” This framework encourages “timely and high-quality geospatial data to support business processes and operations; stronger partnerships across all levels of government and, when appropriate, the private sector, to increase cost efficiency and return on investment; and improve strategies for completing and maintaining nationally significant themes and datasets associated with OMB Circular A-16 to enhance service to citizens” (FGDC, 2010).

****

―Harnessing the Power of Digital Data: Taking the Next Step.‖ Scientific Data Management (SDM) for

Government Agencies: Report from the Workshop to Improve SDM. Workshop held June 29 - July 1, 2010, Washington D.C. March 2011. Report No. CENDI/2011-1. Co-sponsored by the Environmental Protection Agency (EPA), CENDI (The Federal STI Managers Group), and the Interagency Working Group on Digital Data (IWGDD)

**4. University of Oxford Research Data Management Chart**

Found at http://www.admin.ox.ac.uk/rdm/

Research Data Management

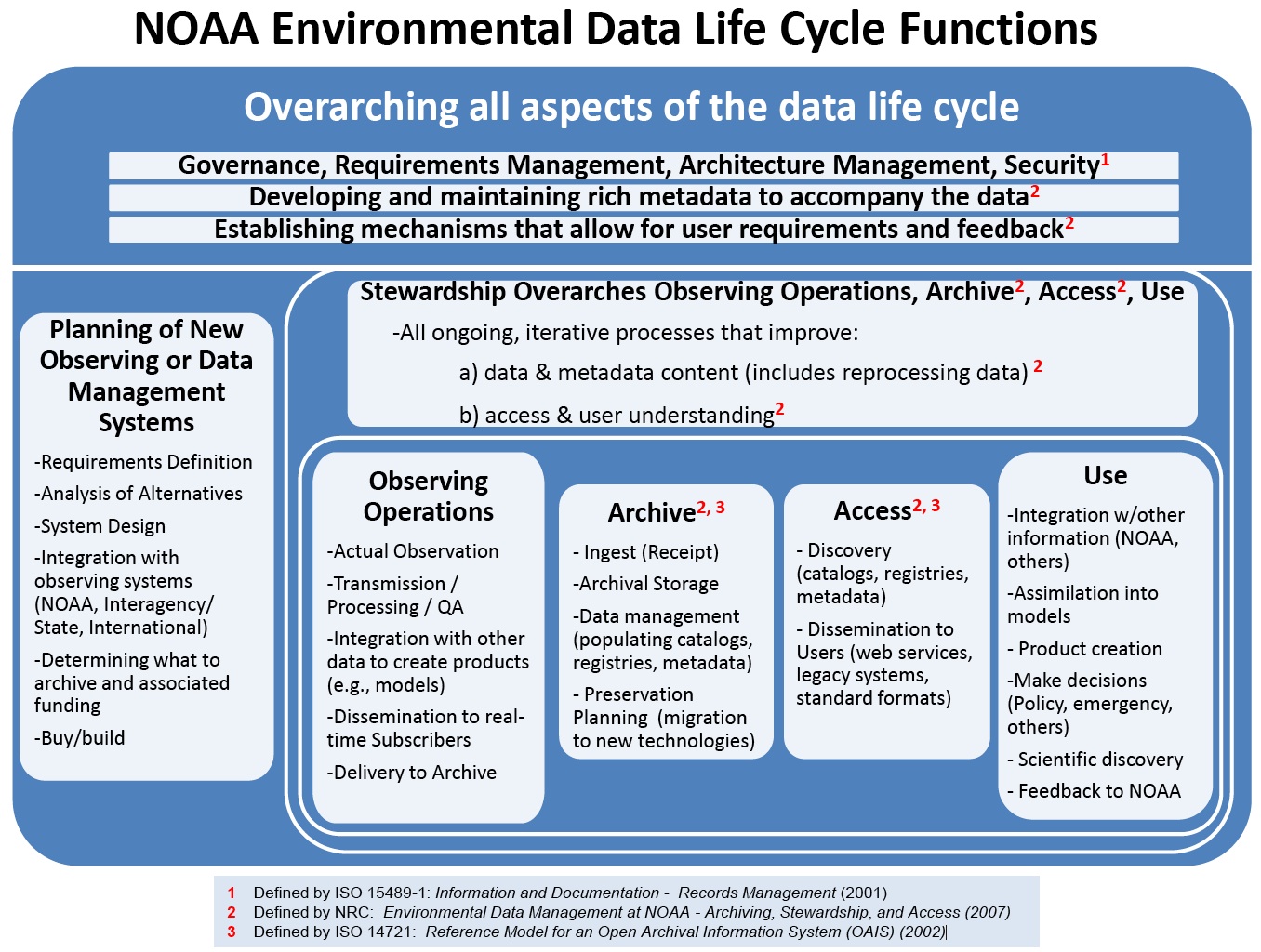
Good practice in data management is one of the core areas of research integrity, or the responsible conduct

of research. The following diagram provides further insight to some of the stages involved in research data management, and the facilities and services available to help, both within the University and from external providers.



**5. NOAA Environmental Data Life Cycle Functions**

Provided by Peter Steurer, NOAA, National Climatic Data Center, 2009.



**6. Open Archival Information System (OAIS) Framework**

Introduction by Larry Baume from NARA. OAIS Reference model found at

http://public.ccsds.org/publications/archive/650x0b1.pdf

OAIS Framework for Developing a “Community of Practice”

Larry Baume

NARA

October 27, 2009

OAIS Framework

• ISO Standard 14721:2002

• Conceptual framework describing the environment, functional components, and

information objects within a system responsible for the long-term preservation

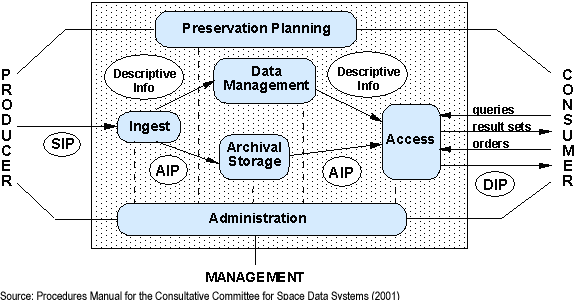
of digital materials

• Lifecycle model for data archives

• Widely recognized in scientific, data management, and archival communities

• Integrates with other ISO standards such at ISO 9000, ISO 15489

OAIS Framework



TRAC Report

• Trusted Repository Audit and Certification methodology, 2007

• OCLC, DPC, and NARA

• Restructured and adopted in Oct. 2009 as CCSDS Recommended Practice

(Red Book)

• Nests with OAIS Framework

• Widely recognized in archival and library communities

CCSDS Red Book

• Audit and certification criteria for measuring “trustworthiness” of a digital

repository

• Lists criteria and evidence needed to support the criteria

• Assumes an external certification process which does not exist now

• Can be used as a self-assessment guide

DRAMBORA

• “Digital Repository Audit Methodology Based on Risk Analysis,” Feb. 2007

• Digital Conservation Centre (DCC) and DigitalPreservationEurope (DPE),

• Proposes a methodological framework, guidelines and audit tools to support the

identification, assessment and management of risks in a digital repository

• Self-assessment checklist, no certification at this time

• Integrates with OAIS framework

Challenges

• No integrated “community of practice” for promoting trusted repositories

• CCDSC Red Book and DRAMBORA are not widely known in the scientific, IT,

and data management communities

• They do not address design of data systems and data management, e.g.

systems planning, funding, system requirements

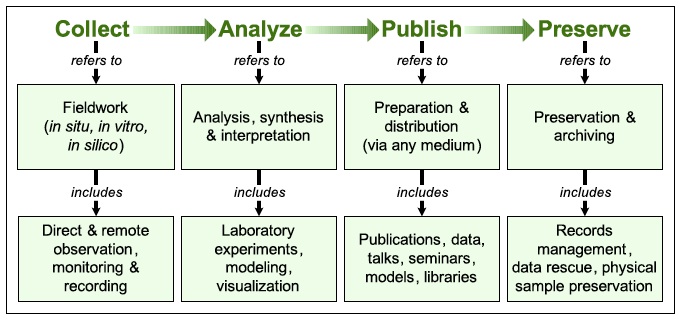
• No Case Studies and Lessons Learned

**7. USGS Scientific Information Management Workshop Vocabulary**

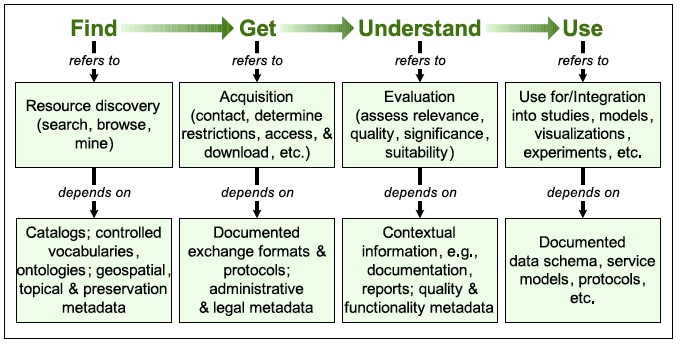
From Tom Gunther and Dave Govoni, USGS, 2006.

**A Vocabulary for Scientific Information and Knowledge Management**

**Producer perspective**



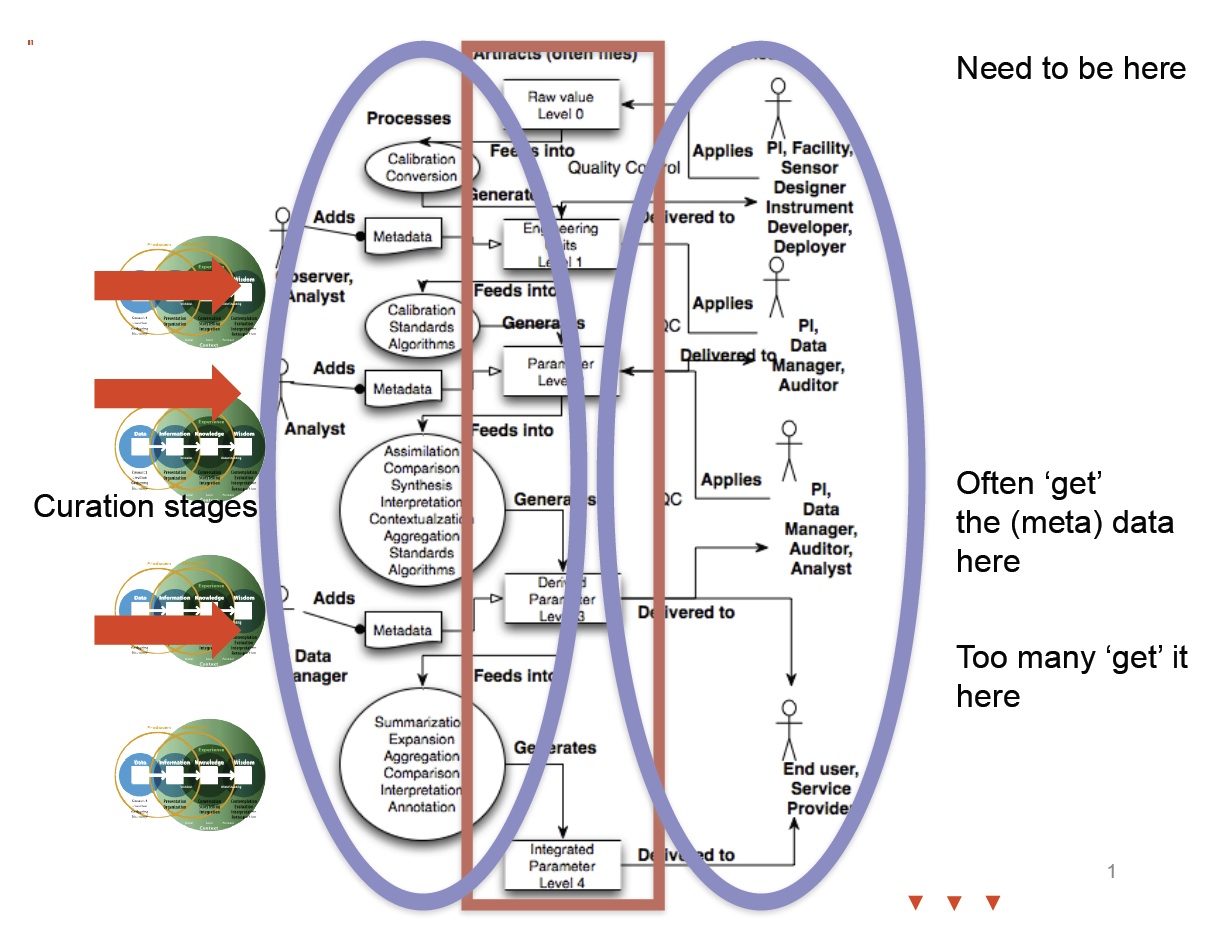
**Consumer perspective**

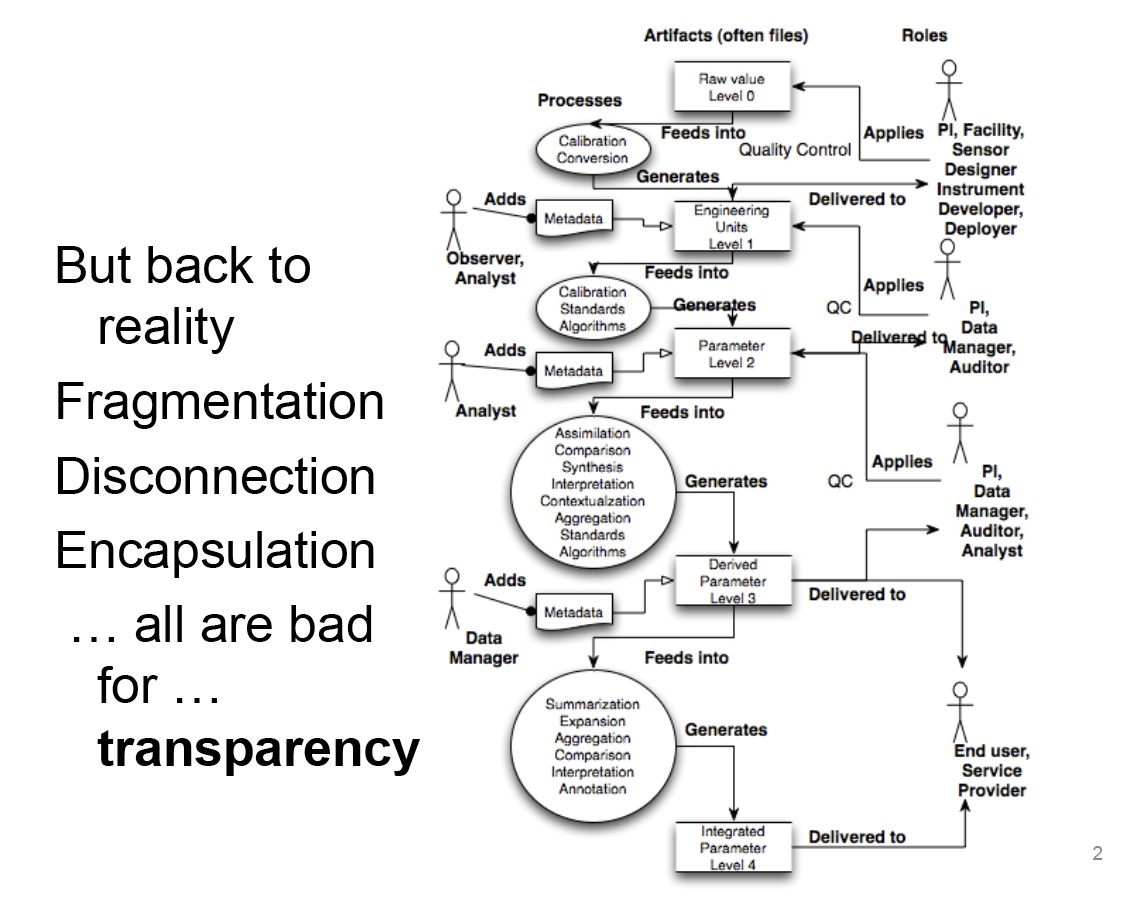


Source: Govoni, D.L. and T.M. Gunther, 2006. Scientific Information Management at the U.S. Geological Survey: Issues, Challenges, and a Collaborative Approach to Identifying and Applying Solutions (Abstract). Geoinformatics 2006—Abstracts. Scientific Investigations Report 2006-5201, p. 19-20. U.S. Geological Survey, Reston, Virginia.

**8. Peter Fox Lifecycle Diagrams**

Presented by Peter Fox, Rennsselaer Polytechnic Institute at USGS CDI August 13, 2010 Denver, CO.





**9. National Science Foundation**

Provided by Vivian Hutchinson (USGS) via email on 16 November 2010.

This is the National Science Foundation's data lifecyle: (taken from the program

solicitations for NSF DataNet projects)...How does it compare to what we are

addressing?

\* Data deposition/acquisition/ingest - Provide systems, tools, procedures, and capacity

for efficient data and metadata deposition by authors and others; acquisition from

appropriate sources; and ingest in accordance with well-developed and transparent

policies and procedures that are responsive to community needs, maximize the potential

for re-use, and ensure preservation and access over a decades timeline.

\* Data curation and metadata management - Provide for appropriate data curation and

indexing, including metadata deposition, acquisition and/or entry and continuing

metadata management for use in search, discovery, analysis, provenance and attribution,

and integration. Develop and maintain transparent policies and procedures for ongoing

collection management, including deaccessioning of data as appropriate.

\* Data protection - Provide systems, tools, policies, and procedures for protecting

legitimate privacy, confidentiality, intellectual property, or other security needs as

appropriate to the data type and use.

\* Data discovery, access, use, and dissemination - Provide systems, tools, procedures,

and capacity for discovery of data by specialist and non-specialist users, access to data

through both graphical and machine interfaces, and dissemination of data in response to

users needs.

\* Data interoperability, standards, and integration - Promote the efficient use and

continuing evolution of existing standards (e.g. ontologies, semantic frameworks, and

knowledge representation strategies). Support community-based efforts to develop new

standards and merge or adapt existing standards. Provide systems, tools, procedures, and

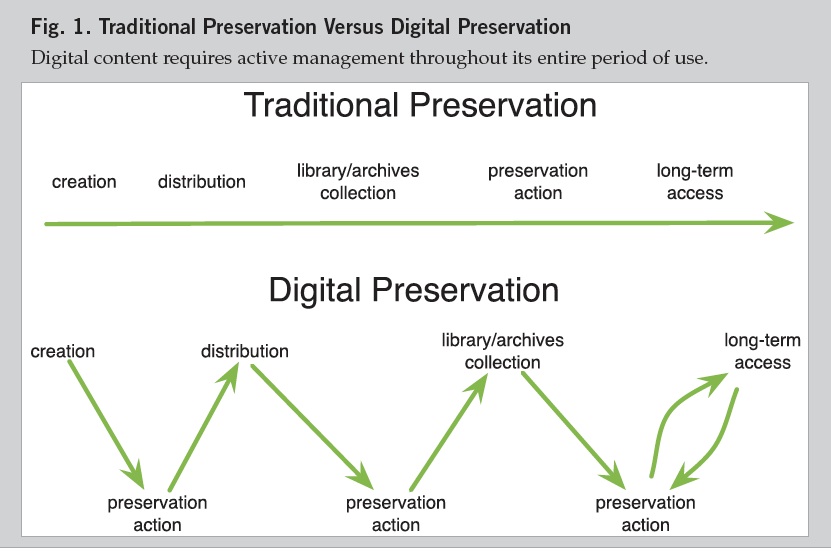
capacity to enhance data interoperability and integration.

\* Data evaluation, analysis, and visualization - Provide systems, tools, procedures, and

capacity to enable data driven visual understanding and integration and to enhance the

ability of diverse users to evaluate, analyze, and visualize data.

**10. NDIIPP Preserving Our Digital Heritage**



Source: National Digital Information Infrastructure & Preservation Program, Preserving Our Digital Heritage: The National Digital Information Infrastructure and Preservation Program 2010 Report, A Collaborative Initiative of the Library of Congress.

**11. What Researchers Want**

A literature study of researchers’ requirements with respect to storage and access to research data.

SURFfoundation

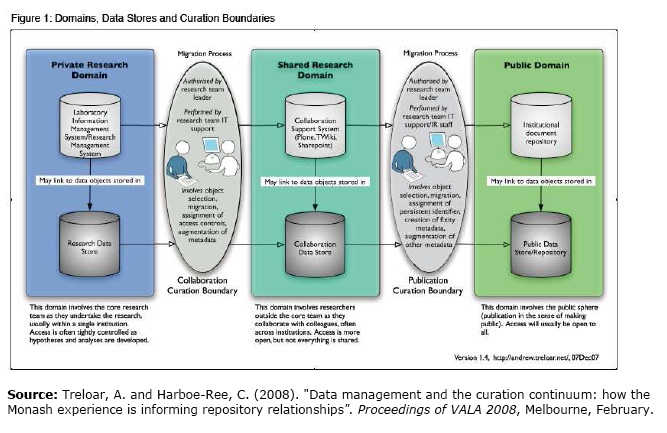
PO Box 2290

NL-3500 GG Utrecht

Author Martin Feijen

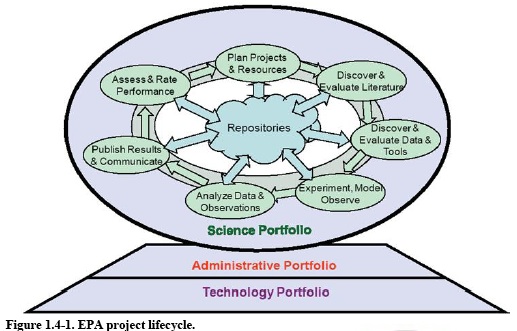
February 2011

Creative Commons Attribution 3.0 Netherlands License.



**12. EPA Project Life Cycle**

U.S. Environmental Protection Agency



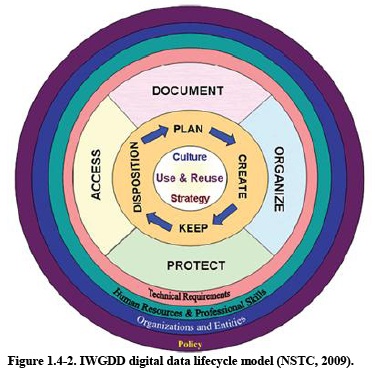
The project lifecycle presented in figure 1.4-1 is a generic model of how science is conducted at its most elemental level. Questions are posed, and projects are planned and resourced to answer those questions. Previous results, data, and publications are reviewed for relevance. Experiments are designed and conducted, and results analyzed and published. The cycle repeats, incorporating lessons learned from previous iterations. Because data often have utility beyond individual projects, the data lifecycle (also presented in figure 1.4-2) [next page], overlaps but differs from the project lifecycle.

―Harnessing the Power of Digital Data: Taking the Next Step.‖ Scientific Data Management (SDM) for

Government Agencies: Report from the Workshop to Improve SDM. Workshop held June 29 - July 1, 2010, Washington D.C. March 2011. Report No. CENDI/2011-1. Co-sponsored by the Environmental Protection Agency (EPA), CENDI (The Federal STI Managers Group), and the Interagency Working Group on Digital Data (IWGDD)

**13. IWGDD’s Digital Data Life Cycle Model**

Interagency Working Group on Digital Data of the Office of Science and Technology Policy [U.S.]



―Harnessing the Power of Digital Data: Taking the Next Step.‖ Scientific Data Management (SDM) for

Government Agencies: Report from the Workshop to Improve SDM. Workshop held June 29 - July 1, 2010, Washington D.C. March 2011. Report No. CENDI/2011-1. Co-sponsored by the Environmental Protection Agency (EPA), CENDI (The Federal STI Managers Group), and the Interagency Working Group on Digital Data (IWGDD)

**14. Scientific Data Management Plan Guidance**

**2.3 Scientific Data should be Managed According to an SDM Plan that Covers the Full Data Lifecycle**

The *Harnessing* guiding principle—―longer preservation, access, and interoperability require management of the full data lifecycle‖—was clearly supported by the planning team‘s observation that both data and project lifecycles are critical. Policy should require that data management planning be well integrated into project planning, as noted in Section 1.4.1. The planning should begin at the inception of the project/effort and should be an integral part of project planning, budgeting, and management. The survey and workshop confirmed this need. In fact, 90% of survey participants declared that after a project begins, a data management plan should be ―an ongoing, open-ended, living document that follows the data through its lifecycle.‖ Discussions also strongly supported the idea that a data management plan is a living document. It is this document and its metadata that connect and document the data lifecycle with the project lifecycle. Survey responses indicate that project data and metadata are both important for context to allow effective secondary use of data by operational and policy users.

Federal science agencies are heterogeneous with respect to their scientific data policy requirements and approaches. Data producers often have little incentive to expend resources on extensive data planning that will primarily benefit secondary users. Research agencies vary from regulatory agencies in this regard. Regulatory agencies use their data to develop regulations, so they must prepare the data for a known secondary use. Legal and scientific defensibility must also be addressed in agency policy.

As agencies increase data management requirements, they must also help create or support the infrastructure needed to allow project managers to fulfill policy requirements. This will ultimately change the entire agency culture and practice regarding scientific data management. For example, if an agency provides an institutional repository to its COPs either directly or through supporting initiatives, then multiple projects could incorporate the repository into their plans for storage and archiving. This will help the project manager to define this part of the lifecycle and will help to realize economies of scale and coordination.

As the government looks to its plans for open government through the development of tools such as Data.gov, it is important to integrate these tools into the overall federal architecture and project lifecycle. Federal objectives for transparency and open access to data can only be met sustainably and economically if they are (1) integrated into the business process of science and (2) supported by an interoperable federal architecture. SDM policies and planning are needed to enable this environment to exist.

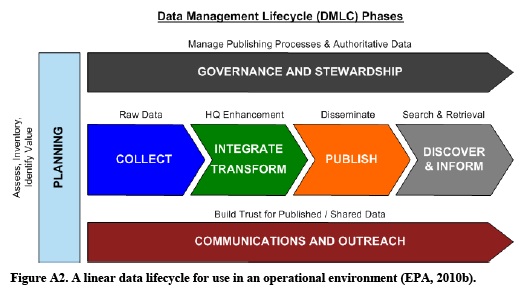
At the bottom line, agency policy should acknowledge the importance of the project lifecycle context in the data management lifecycle to facilitate data reuse.

―Harnessing the Power of Digital Data: Taking the Next Step.‖ Scientific Data Management (SDM) for

Government Agencies: Report from the Workshop to Improve SDM. Workshop held June 29 - July 1, 2010, Washington D.C. March 2011. Report No. CENDI/2011-1. Co-sponsored by the Environmental Protection Agency (EPA), CENDI (The Federal STI Managers Group), and the Interagency Working Group on Digital Data (IWGDD)

**15. Linear Data Life Cycle**

A linear lifecycle model was suggested as being easier to work with in an operational environment. Figure A2 shows a linear lifecycle adapted for use. This model was developed by the Environmental Protection Agency (EPA) Office of Research and Development (ORD). The figure highlights the importance of governance and communications as key aspects of implementing a lifecycle approach. Element processes are defined below.



1. **1. Plan:** Data are assessed and inventoried by open government governance bodies and segment architects, and high-value sets are identified for sharing with the public through open government initiatives such as Data.gov.
2. **2. Collect:** Data are collected by the source entity, source providers push data to federal source systems, which provide data in a specified file format, and then deltas are managed by source record tags in the specified file format.
3. **3. Integrate and Transform:** The agencies integrate the raw data that were collected and add value through various means, including input from individual program offices and scientific research projects. The data are transformed from their initial state and stored in a value-added state, such as through web services.
4. **4. Publish:** Information resources are prepared for publishing to one or more of the many audiences, including congress, the public, tribal governments, academia, research and scientific partners in non-governmental organizations, other federal agencies, and other stakeholders (e.g., industry, communities, researchers, the media, and Data.gov audiences).
5. **5. Discovery:** Agencies manage search and retrieval for various internal and external audiences. Discovery will become more complex as secondary audiences are supported through open government initiatives. Secondary audiences need to be informed of the meaning of data as understood by primary audiences, who are more familiar with the environmental legal landscape.
6. **6. Governance and Stewardship:** This element of the process defines governance bodies and agendas, and it gains acceptance of data steward roles. Governance and stewardship manage the publishing process for ongoing change control, and they maintain versions of the truth across common data.
7. **7. Communications and Outreach:** This aspect provides for inventory of high value data sets, enables technologies and controlled vocabularies for re-use, allows for management of information exchange agreements, and encourage re-use through communications.

―Harnessing the Power of Digital Data: Taking the Next Step.‖ Scientific Data Management (SDM) for

Government Agencies: Report from the Workshop to Improve SDM. Workshop held June 29 - July 1, 2010, Washington D.C. March 2011. Report No. CENDI/2011-1. Co-sponsored by the Environmental Protection Agency (EPA), CENDI (The Federal STI Managers Group), and the Interagency Working Group on Digital Data (IWGDD)

**16. Generic Science Data Lifecycle**

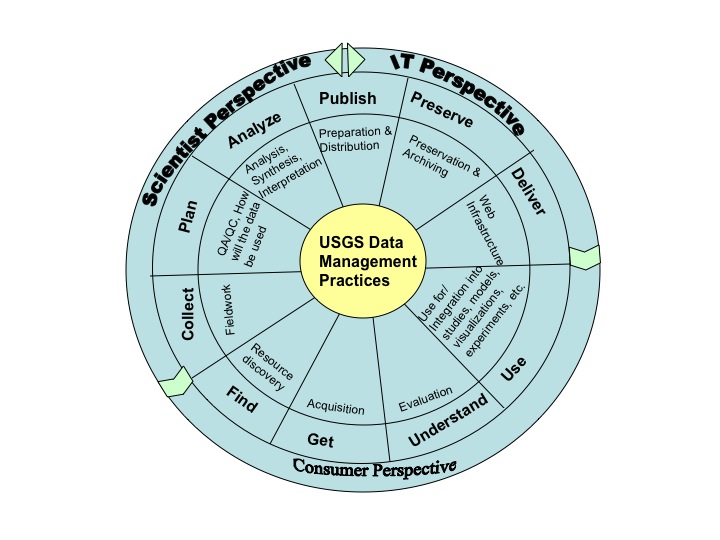


The science data lifecycle is given in Figure C2. The data outputs of a science project are freestanding artifacts that must be maintained intact, secure, and accessible for future uses, foreseen and unforeseen, but archived or disposed of when no longer useful. Within this linear construct, many data maintenance feedback loops may be interpreted through this simplified model. Programs will plan next set of data acquisitions based on discoveries from the current one and would use lessons learned from data management to plan the evolution of data system for future datasets.

―Harnessing the Power of Digital Data: Taking the Next Step.‖ Scientific Data Management (SDM) for

Government Agencies: Report from the Workshop to Improve SDM. Workshop held June 29 - July 1, 2010, Washington D.C. March 2011. Report No. CENDI/2011-1. Co-sponsored by the Environmental Protection Agency (EPA), CENDI (The Federal STI Managers Group), and the Interagency Working Group on Digital Data (IWGDD)

**17. Cassandra Ladino Hybrid Data Lifecycle Model**



**18. Ray Obuch Data Management – A Lifecycle Approach**

**USGS Information Management Workshop (May 1997)**

**What is The Life Cycle Approach to Data Management?**

**The Data Management Life Cycle Phases**

**1) Creation**

**2) Dissemination**

**3) Access**

**4) Use**

**5) Preservation**

**6) Evaluation**

**Phase 1**

***Creation***

***Production Agency Perspective***

**Create Government information in a variety of useful formats and in**

**consultation with other program partners.**

**Comply with 17 USC 105**

**17 USC Sec. 105. - Subject matter of copyright: United States Government works**

Copyright protection under this title is not available for any work of the United States Government, but the United States Government is not precluded from receiving and holding copyrights transferred to it by assignment, bequest, or otherwise

**Phase 1**

***Creation***

***Central Operational Authority Perspective***

**Facilitate communication between Program partners in the design and development of information products and services**

**Phase 1**

***Creation***

***Participating Libraries Perspective***

**As intermediaries, cooperate with Program partners in the design and development of information products and services and facilitate user feedback.**

**Phase 1**

***Creation***

***End User Perspective***

**As primary clientele, cooperate with Program partners in the design and development of information products and services.**

**Phase 2**

***Dissemination***

***Producing Government Agencies Perspective***

**Provide government information products and services through multi faceted dissemination programs at no cost to the public through participating libraries.**

**Phase 2**

***Dissemination***

***Central Operational Authority Perspective***

**Distribute or coordinate the distribution of products and services in a timely fashion.**

**Provide a variety of dissemination options and channels.**

**Phase 2**

***Dissemination***

***Participating Libraries Perspective***

**Work with other program partners to ensure the timely dissemination of government information through a variety of dissemination programs***.*

**Phase 2**

***Dissemination***

***End User Perspective***

**Work with other Program partners to require that government information is being disseminated through a variety of channels and that it is appropriate to their needs.**

**Phase 3**

***Access***

***Producing Government Agencies Perspective***

**Release products and services in a timely and useable fashion.**

**Notify Program partners through the Central Authority about existing, planned, changing, or discontinued products and services**

**Develop GILS and other locator systems to help identify government information products and services.**

**Phase 3**

***Access***

***Central Operational Authority Perspective***

**Identify, obtain, or provide access to government information products and services regardless of format.**

**Develop catalogs, pathfinders, and other locator systems to identify government information products and services.**

**Establish standards and enforce regulations that ensure Program**

**compliance**

**In sales program, charge no more than marginal cost of dissemination.**

**Phase 3**

***Access***

***Participating Libraries Perspective***

**Provide timely access to government information at not fee to the user regardless of their geographic location or ability to pay.**

**Share resources and expertise through interlibrary loan, document delivery, reference assistance, and electronic networks.**

**Supplement distributed Program products with commercially produced indexes, publications and equipment necessary to meet**

**public needs.**

**Phase 3**

***Access***

***End User Perspective***

**Own publicly supported government information products and services and therefore must always have Guaranteed access to them.**

**Phase 4**

***Use***

***Producing Government Agencies Perspective***

**Provide documentation, software, technical support and user training.**

**Phase 4**

***Use***

***Central Operational Authority Perspective***

**Distribute/coordinate access to government information to program partners a no charge.**

**Phase 4**

***Use***

***Participating Libraries Perspective***

**As intermediaries, assist users in the identification, Location, use and acquisition of government information Regardless of format.**

**Phase 4**

***Use***

***End User Perspective***

**Government information products and services must Always be provided in usable format to the public.**

**Phase 5**

***Preservation***

***Producing Government Agencies Perspective***

**Cooperate with other Program participants to ensure That all information products are archived, accessible, accurate, and compatible with current and future technologies.**

**Phase 5**

***Preservation***

***Central Operational Authority Perspective***

**Ensure that all information products are archived, accessible, accurate, and compatible with current and future technologies.**

**Phase 5**

***Preservation***

***Participating Libraries Perspective***

**Cooperate with other Program participants to ensure That all information products are archived, accessible, accurate and compatible with current and future technologies.**

**Phase 5**

***Preservation***

***End User Perspective***

**Must always have access to government information in well-preserved, accessible, and accurate condition.**

**Phase 6**

***Evaluation***

***Producing Government Agencies Perspective***

**Solicit and consider input from Program partners in the evaluation of government information products and services.**

**Phase 6**

***Evaluation***

***Central Operational Authority Perspective***

**Provide avenues for the evaluation of the Program Including advisory councils, Federal agencies, libraries, and the general public.**

**Phase 6**

***Evaluation***

***Participating Libraries Perspective***

**Work with other Program partners to determine the success of the Program through formal and informal evaluations.**

**Phase 6**

***Evaluation***

***End User Perspective***

**Establish criteria and provide through formal and informal evaluation the necessary feedback to determine the success of the Program.**

**Conclusions**

**Adopting the “life cycle” approach and evaluating work flow and data flow throughout the organization will:**

• **Facilitate communication and planning**

• **Improve current data management**

• **Help meet the demand for digital products & data**

• **Enhance return on investment by:**

• **Accomplishing tasks on time and within budget**

• **Re-utilization of data and information**

**19. USGS Data Management Plan Framework (DMPf) – Smith, Tessler, and McHale, 2010 [Climate Effects Network (CEN) and Alaska Science Center (ASC)]**

The DMPf recognizes three distinct phases in research-based data management:

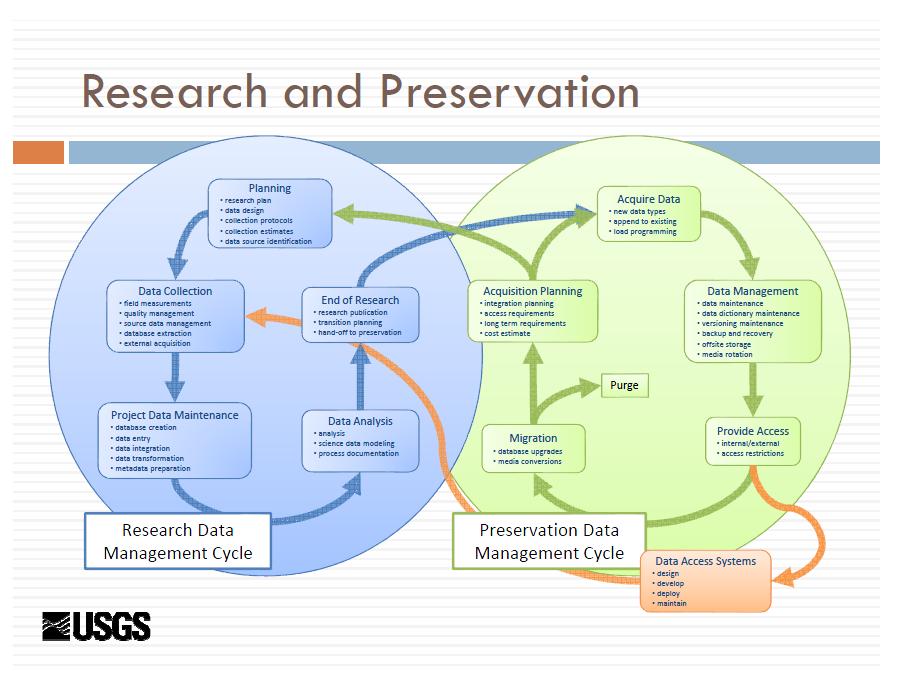
(1) Research Data Management

(2) Preservation Data Management

(3) Data Exposure and Delivery

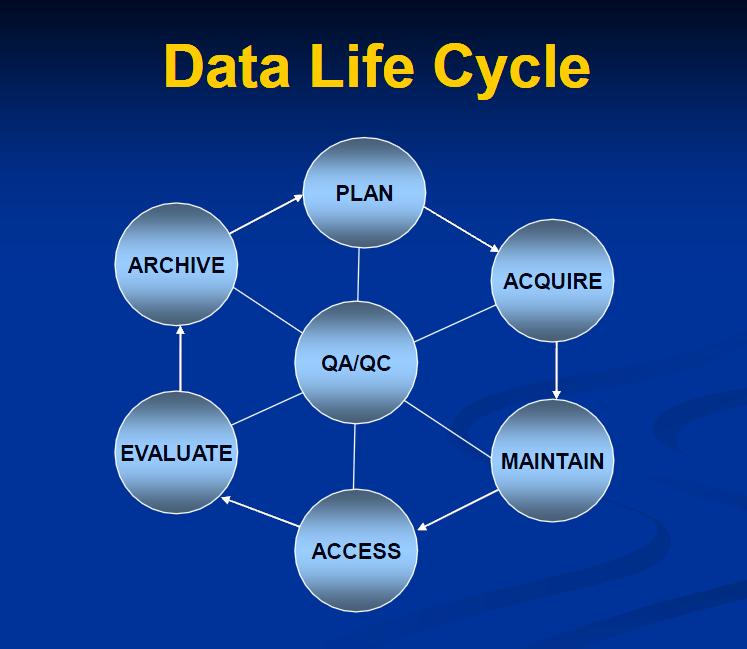
Each of these phases involve data management activities but they have different overall objectives, utilize different people, roles, and skills, and usually imply different funding, physical infrastructure, and technical requirements and support functions. All share the goal of providing for the quality, provenance, and contextual integrity of the data, yet there is a glaring need for some level of coordination, such as research requirements that facilitate preservation, and preservation metadata and formats that facilitate dataset cataloging and delivery. The missing connection simply highlights our current lack of enterprise-level data management standards (either DOI or USGS) and an overall data-integration strategy.

In particular it is important to recognize that the Research Data Management Cycle is separate from the Preservation Data Management Cycle, although the two are often discussed simultaneously, and often place an unacceptable ‘data management’ burden on the research team when they are not preservation specialists. Data-sharing through Exposure and Delivery activities after the conclusion of the active research phase is often viewed as an ‘end point’ separate from the others, yet the best data source for delivery to users will be obtained from those well-preserved and documented data stores (preservation archives) that resulted from well-documented data developed during the active research phase. These pieces of a full DMPf are separate and interrelated, but an effective data management strategy requires that they be interdependent as well.

***Many opportunities for confusion, conflict, and working at cross-purposes in data management activities can occur when the separate nature of the three DMPf phases are not recognized and acknowledged.***

**20. BLM Data Management Handbook (and related materials)**

* The Data Life Cycle is designed to provide a framework for data management.
* That framework is intended to allow for management of the data independent of the system or application that it resides in.
* The Data Life Cycle has been intentionally drawn to be ‘non-linear’.
* While there is a logical pattern to the life cycle, where necessary, the user may move in-between stages as needed.
* It is also important to note that QA/QC is located in the center where it can touch on all stages of the life cycle.



All the questions of documentation, storage, quality assurance, and ownership then need to be answered for each stage of the data life cycle, starting with the recognition of a need, and ending with archiving or updating the information.

**QA/QC involved with each stage are:**

* **Plan**: Decide on the quality level, method for measuring quality and how often the quality should be evaluated.
* **Acquire**: Establish the data acquisition acceptance criteria and the acceptance testing process.
* **Maintain**: Conduct on-going data improvement.
* **Access**: Review and confirm access permissions.
* **Evaluate**: Evaluate quality control procedures, acceptable quality levels, and quality control results.
* **Archive**: Ensure metadata file is current and is archived with the actual data.

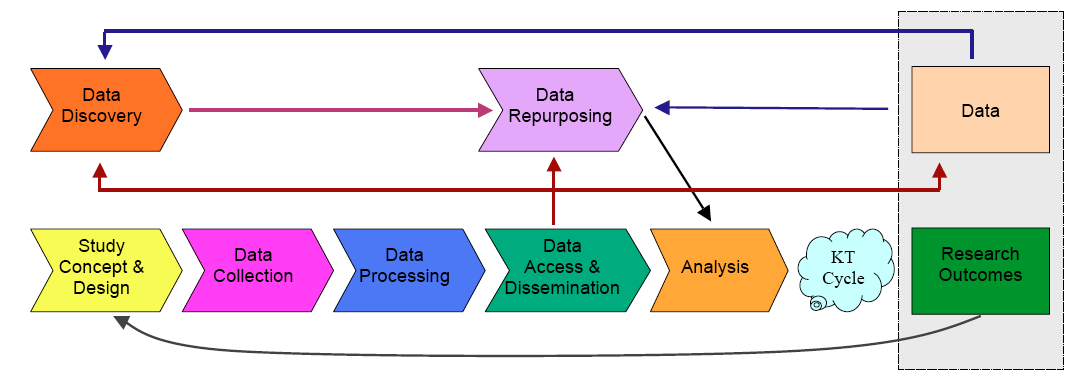
**21.** **ARL Joint Task Force on Library Support for E-Science**

*Final Report and Recommendations to the Scholarly Communication Steering Committee, the Public Policies Affecting Research Libraries Steering Committee, and*

*the Research, Teaching, and Learning Steering Committee*

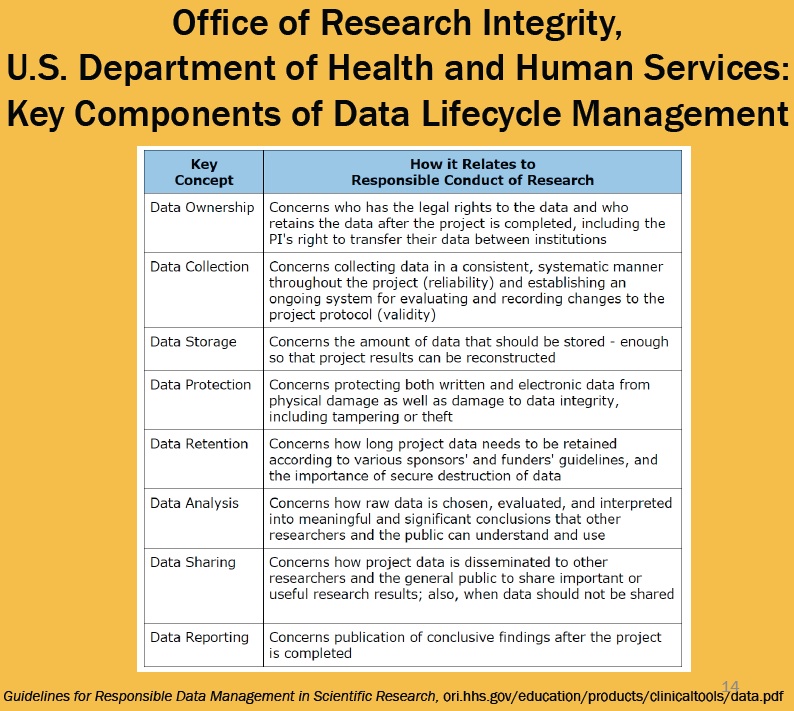
(http://www.arl.org/bm~doc/ARL\_EScience\_final.pdf)

From the perspective of the life cycle of research data, preservation occurs through the stages of data production and the creation of research outputs represented on the bottom row. Long-term preservation in this model consists of the practices followed in caring for the data, which is represented by the box on the right side of the figure. Data curation is characterized on the top row by the stages of data discovery and data repurposing, which make use of the preserved data. The activities of these two functions bring new value to the collection through analyses of the metadata, which display aspects of the collection in new light, and the creation of new data from the existing data collection.



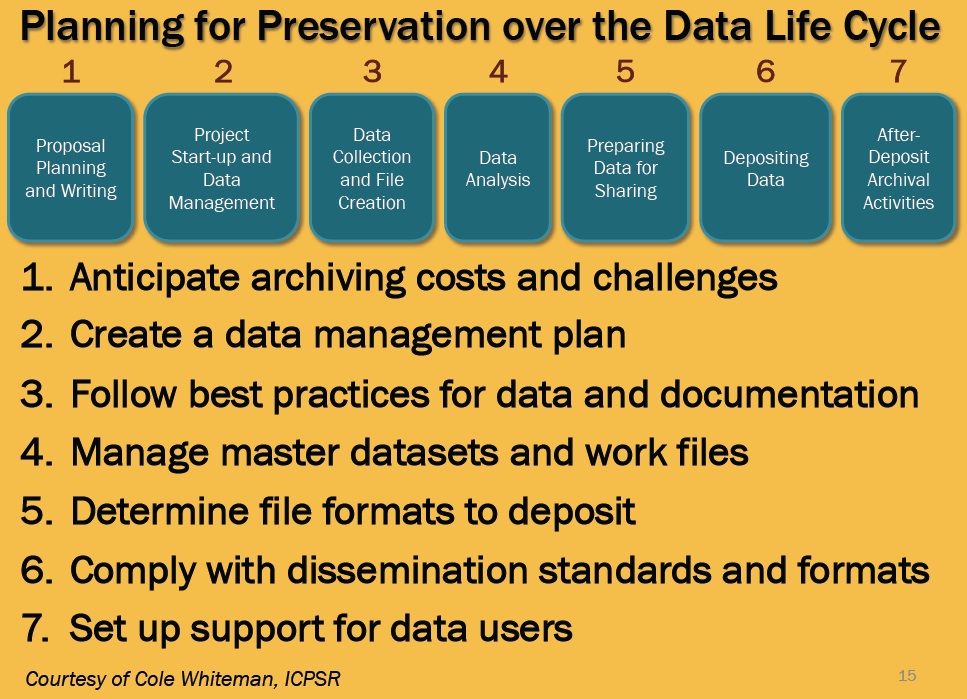
The “KT Cycle” in the diagram represents the processes of knowledge transfer. This life cycle diagram comes from Charles Humphrey, “E-Science and the Life Cycle of Research” (2006) available online at http://datalib.library.ualberta.ca/~humphrey/lifecycle-science060308.doc.

**22. U.S. Department of Health and Human Services Key Components**



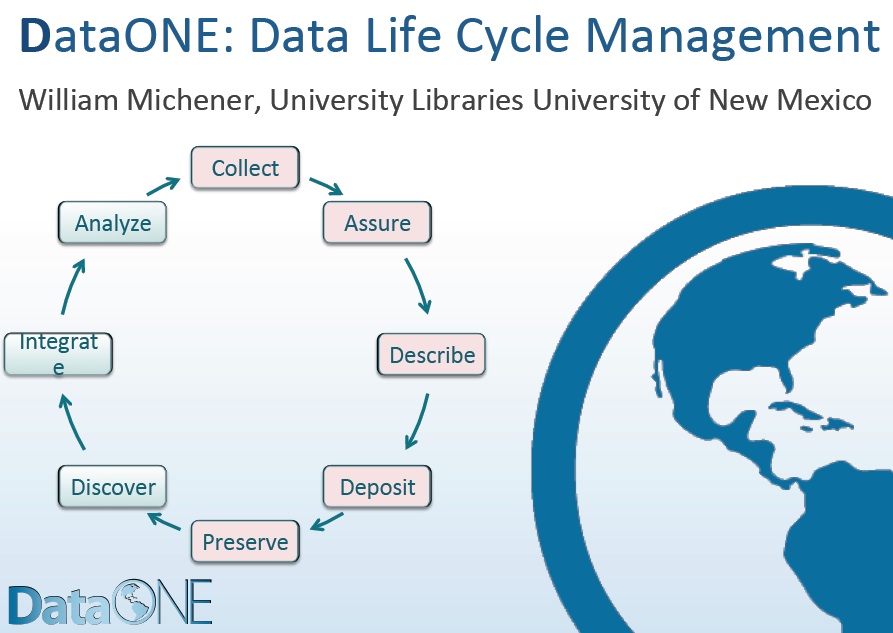
Presented by Jose-Marie Griffiths, Bryant University, at the National Science Foundation Research Data Lifecycle Management Workshop Princeton, NJ July 18-20, 2011.

**23. Interuniversity Consortium for Political and Social Research (ICPSR) Preservation over the Data Life Cycle**



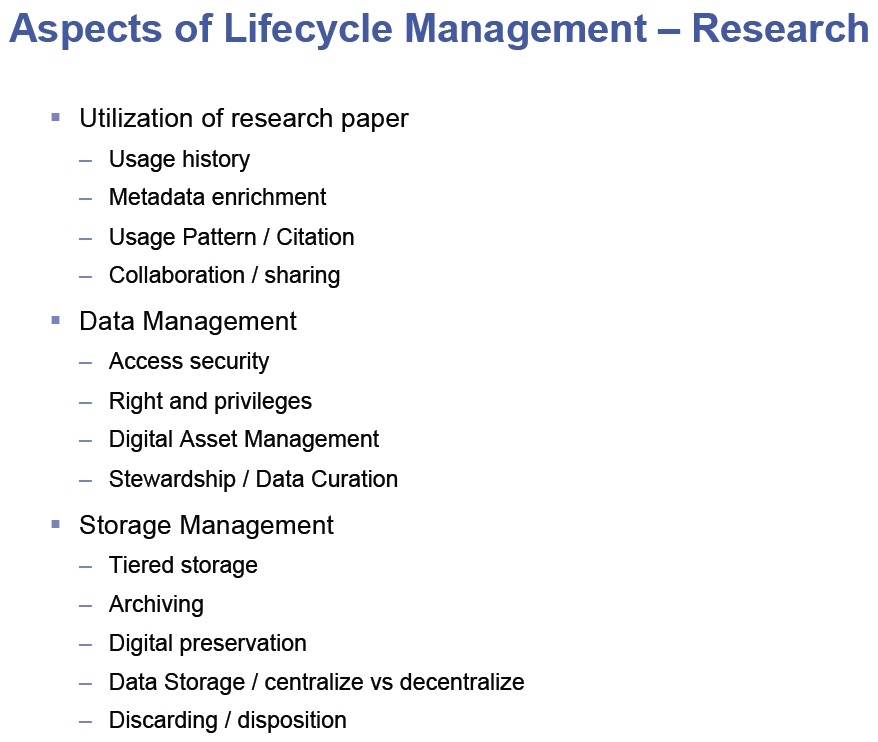
Presented by Jose-Marie Griffiths, Bryant University, at the National Science Foundation Research Data Lifecycle Management Workshop Princeton, NJ July 18-20, 2011.

**24. William Michener DataONE: Data Life Cycle Management**



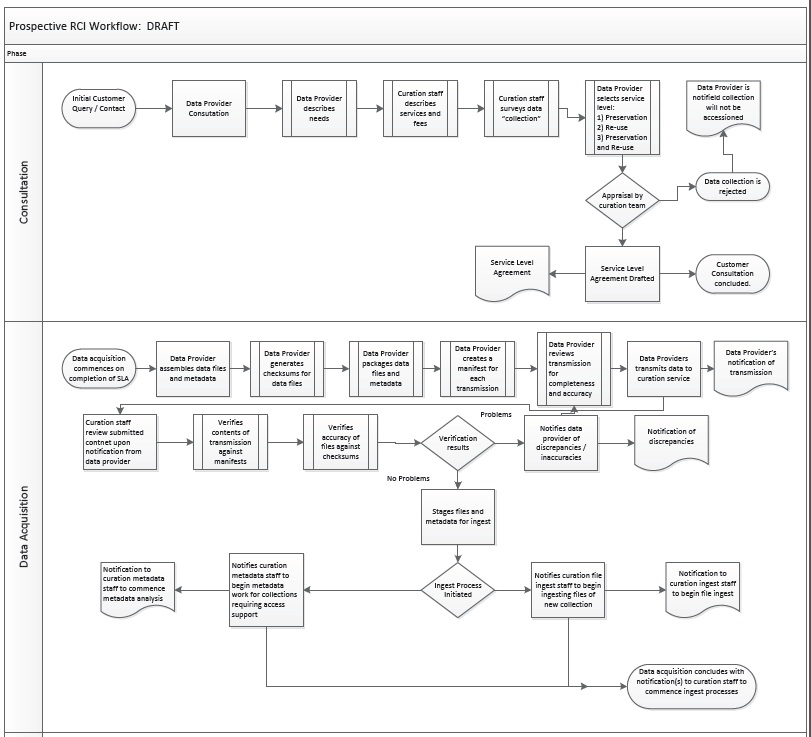
Presented by William Michener, University of New Mexico, at the National Science Foundation Research Data Lifecycle Management Workshop Princeton, NJ July 18-20, 2011.

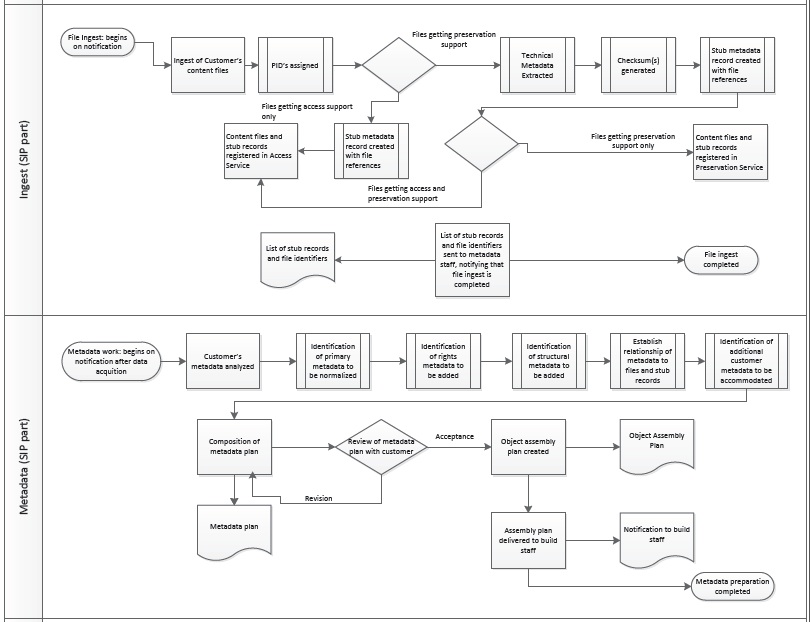
**25. IBM Aspects of Lifecycle Management – Research**

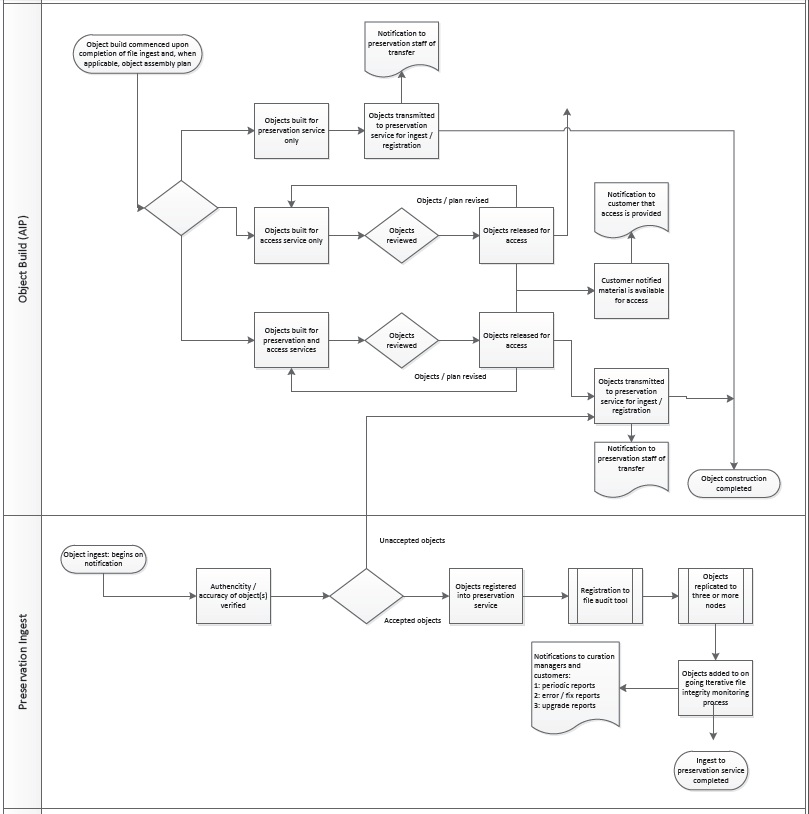


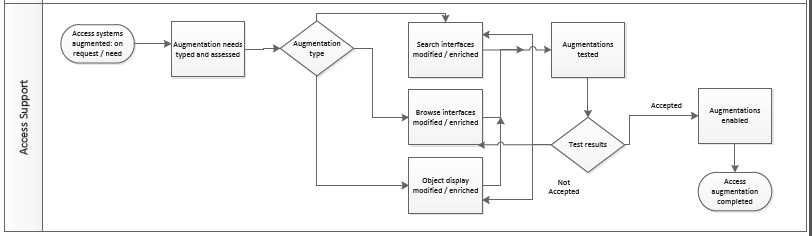
Presented by Imtiaz Khan, IBM, at the National Science Foundation Research Data Lifecycle Management Workshop Princeton, NJ July 18-20, 2011.

**26. University of California San Diego Digital Curation Program**



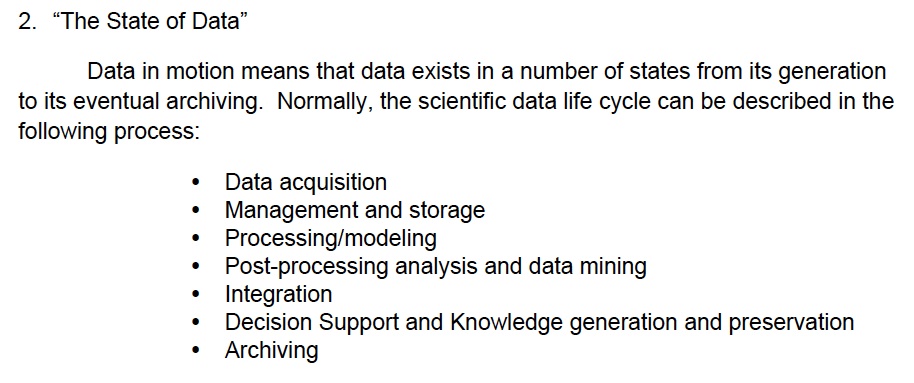






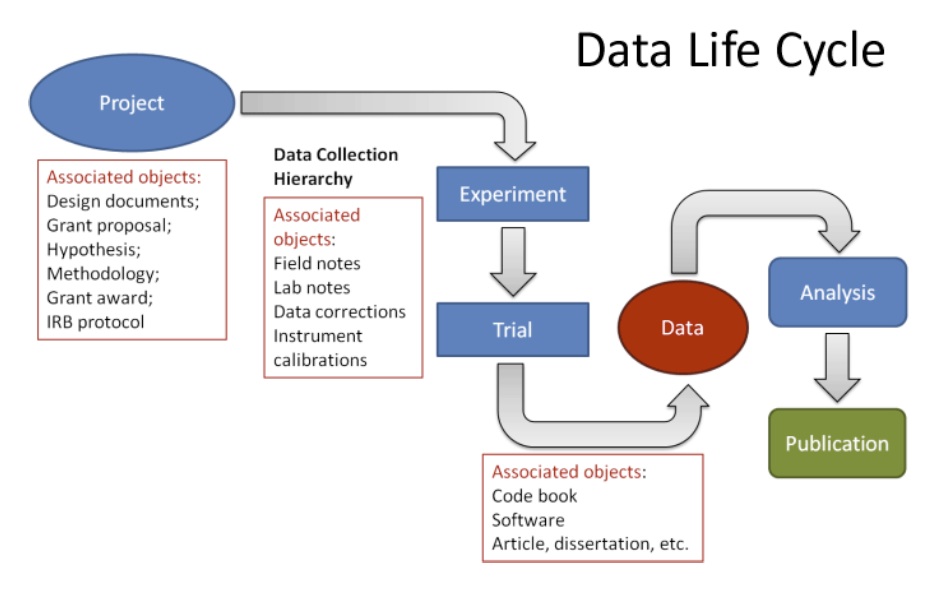
Submitted by David Minor, University of California San Diego Libraries and San Diego Supercomputer Center, Declan Fleming, Ardys Kozbial, and Brad Westbrook, University of California San Diego Libraries to the National Science Foundation Research Data Lifecycle Management Workshop Princeton, NJ July 18-20, 2011.

**27. University of Miami Scientific Data Lifecycle**



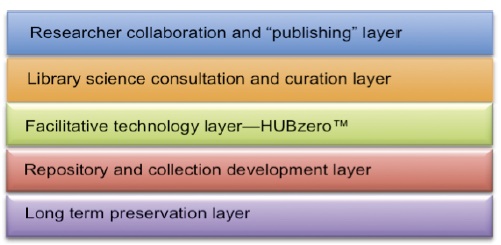
Submitted by Nicholas F. Tsinoremas, Joel Zysman, Christopher Mader, and Jay Blaire, Center for Computational Science, Universtity of Miami, to the National Science Foundation Research Data Lifecycle Management Workshop Princeton, NJ July 18-20, 2011.

**28. Managing Research Data Lifecycles through Context**



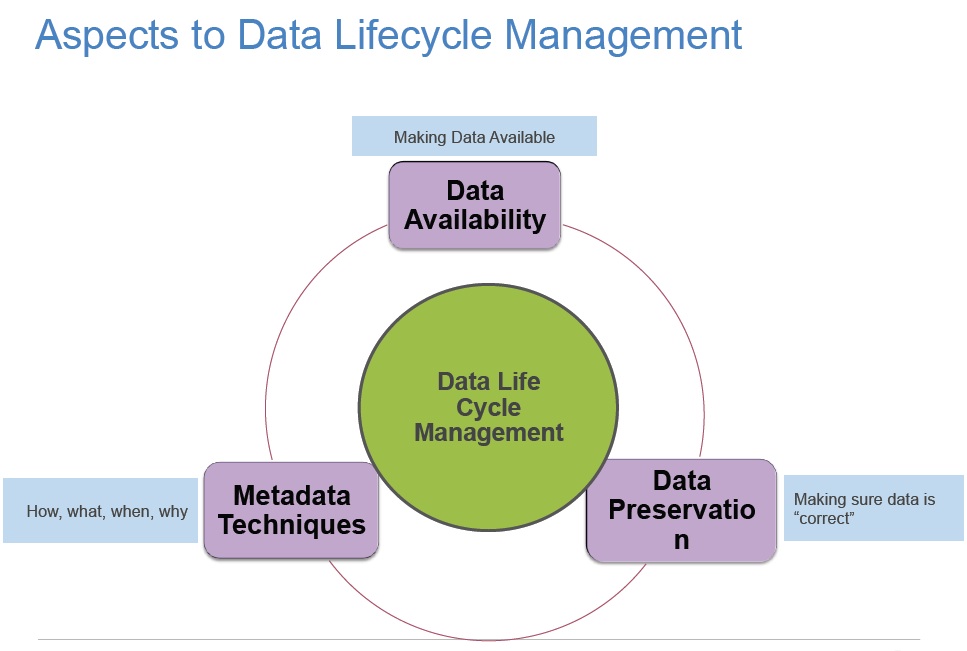
Submitted by Grace Agnew and Ryan Womack, Rutgers University Library, to the National Science Foundation Research Data Lifecycle Management Workshop Princeton, NJ July 18-20, 2011.

**29. Purdue University Research Repository Collaborative Model**



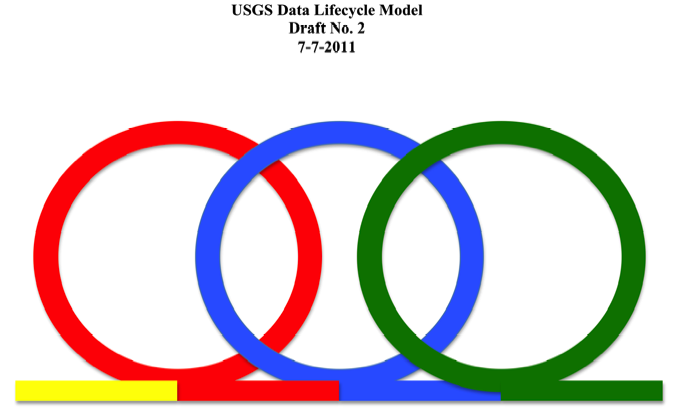
Submitted by D. Scott Brandt, Purdue University Libraries, to the National Science Foundation Research Data Lifecycle Management Workshop Princeton, NJ July 18-20, 2011.

**30. Data Lifecycle Management at Dell**



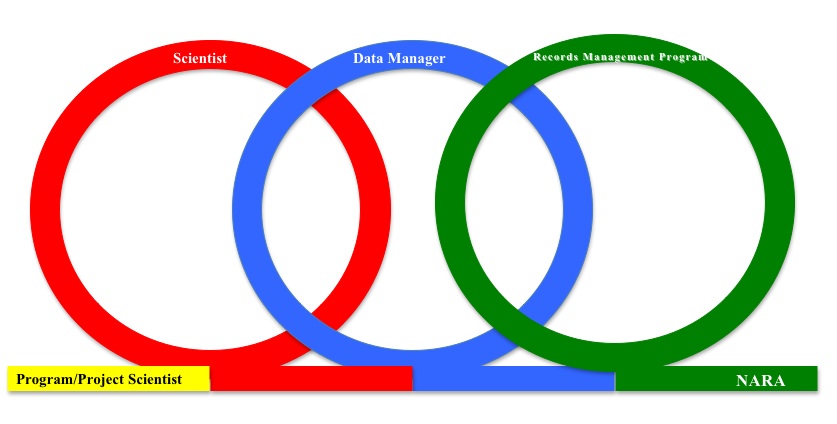
Presented by Jeffrey Layton, Dell, at the National Science Foundation Research Data Lifecycle Management Workshop Princeton, NJ July 18-20, 2011.

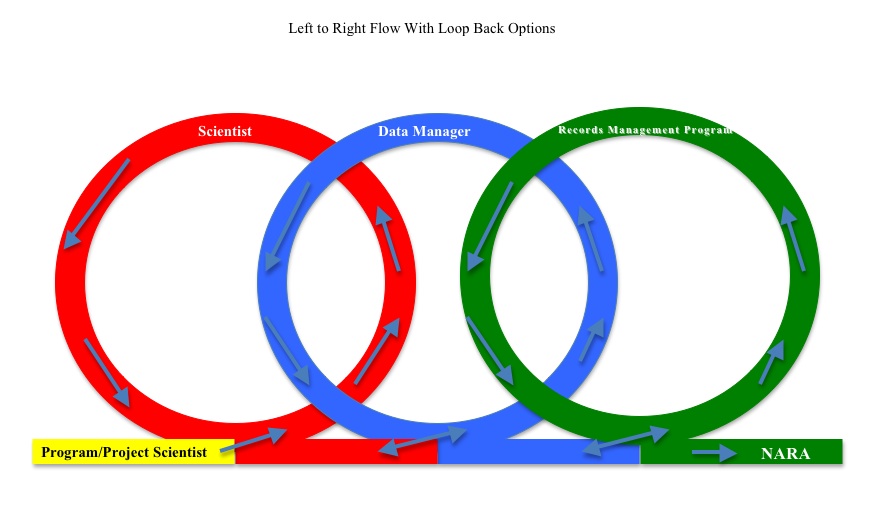
**31.** **John Faundeen & Ellyn Montgomery “Spins”**

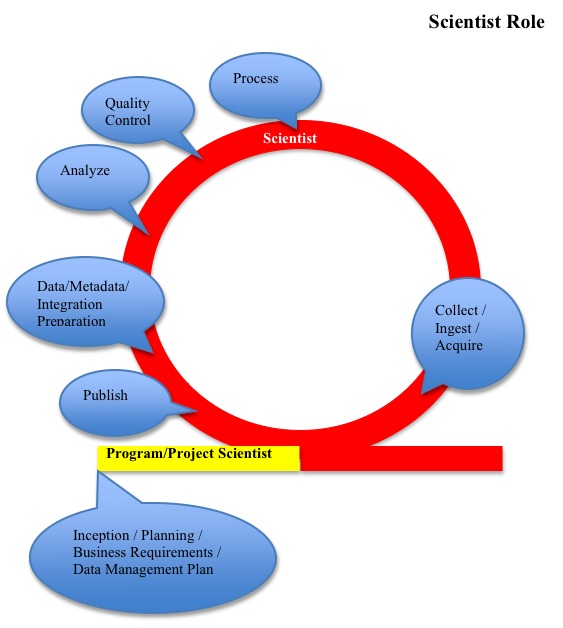


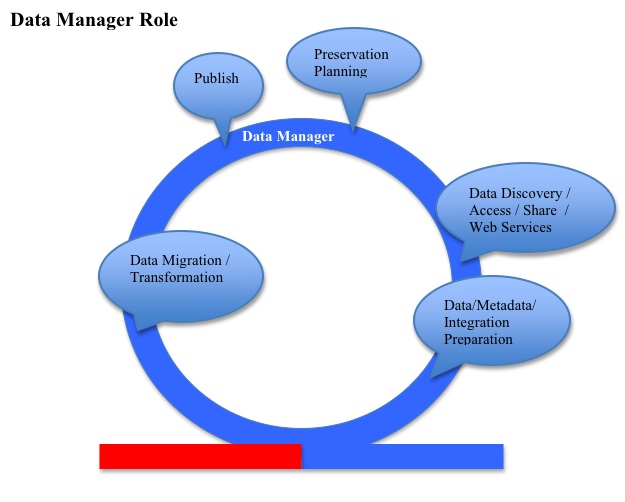
**Roles and Timeline**

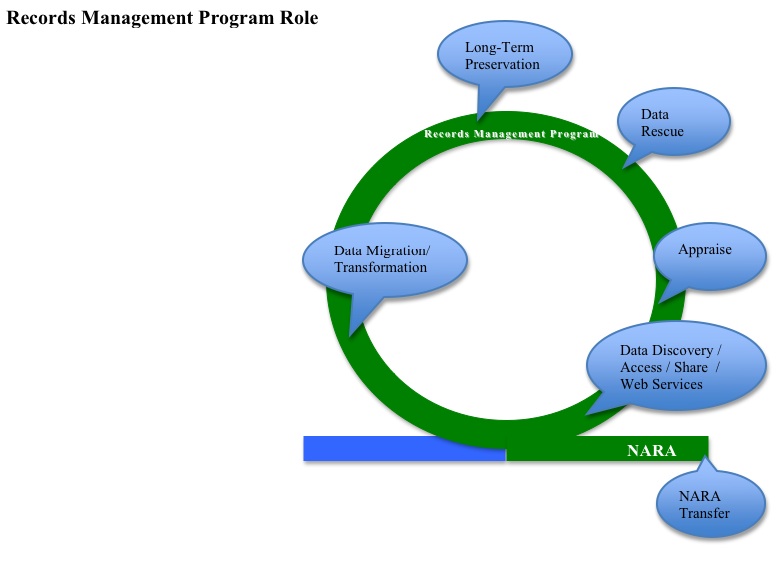
(left to right reading)

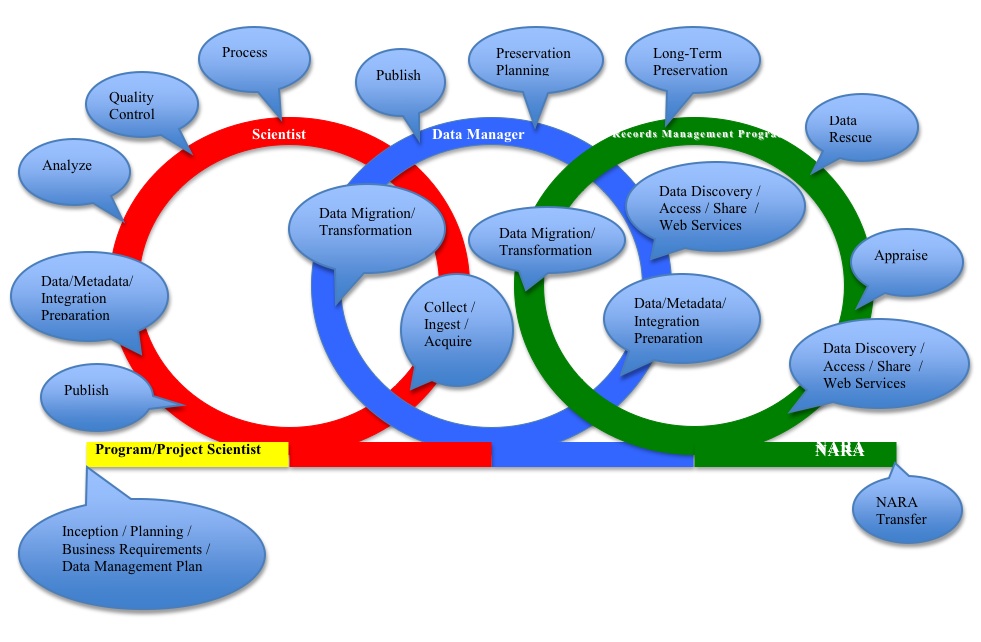


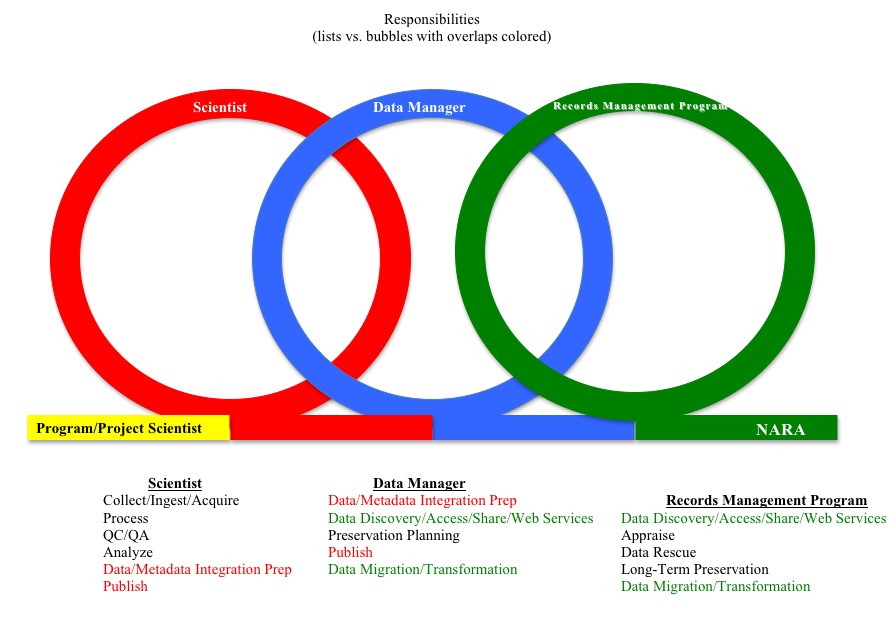


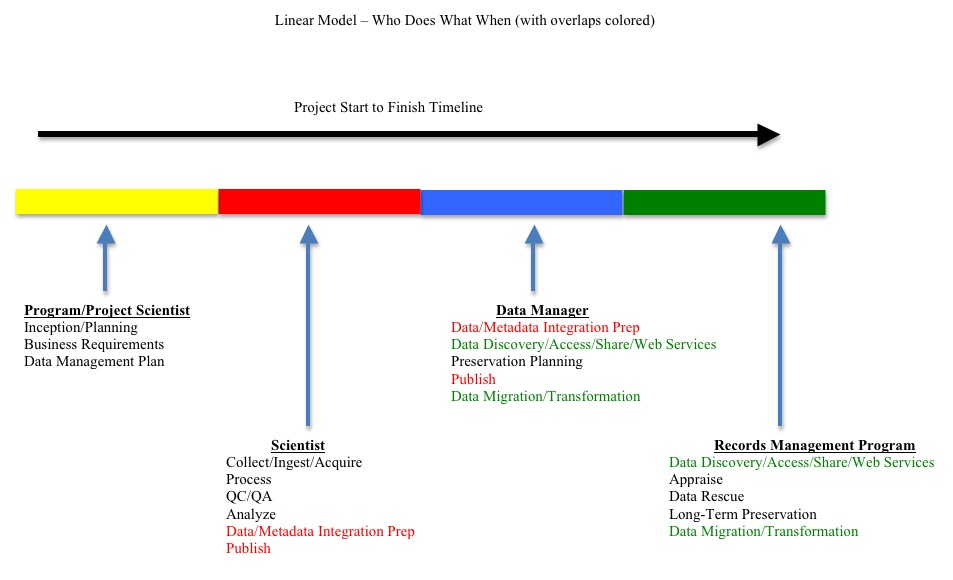


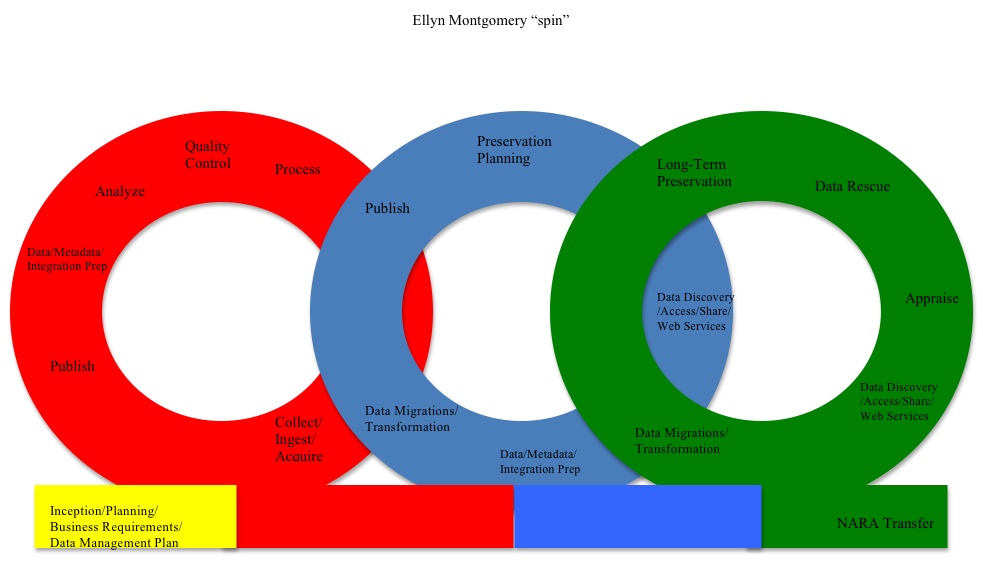




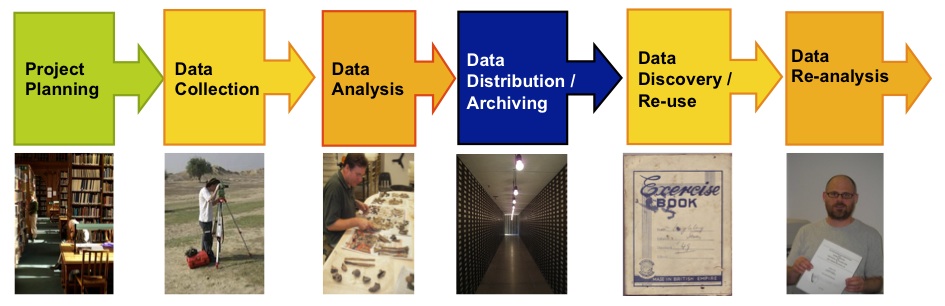








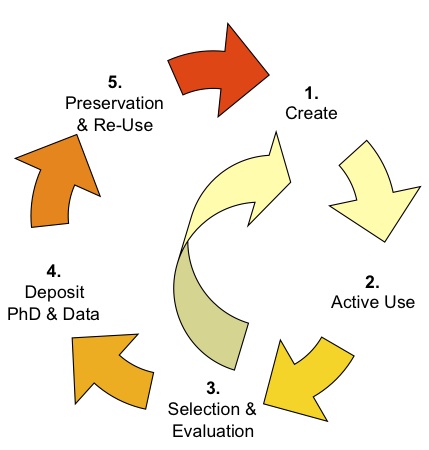
**32. DataTrain Model 1**





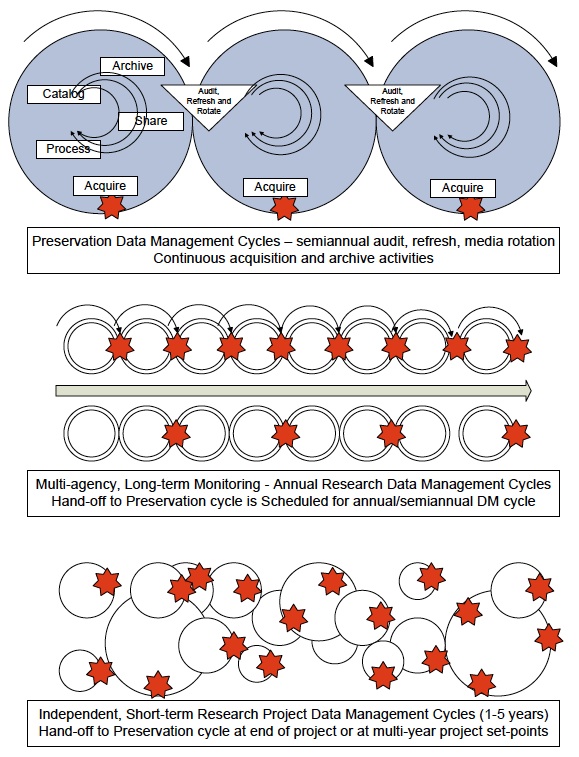
**33. DataTrain Model 2**

**Scientist**

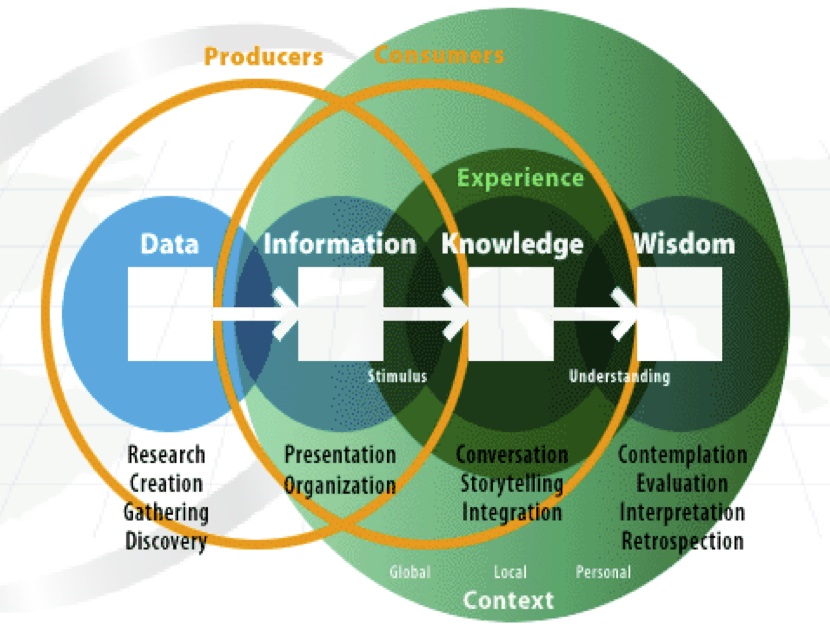




1. **Steve Tessler Data Management Cycles**

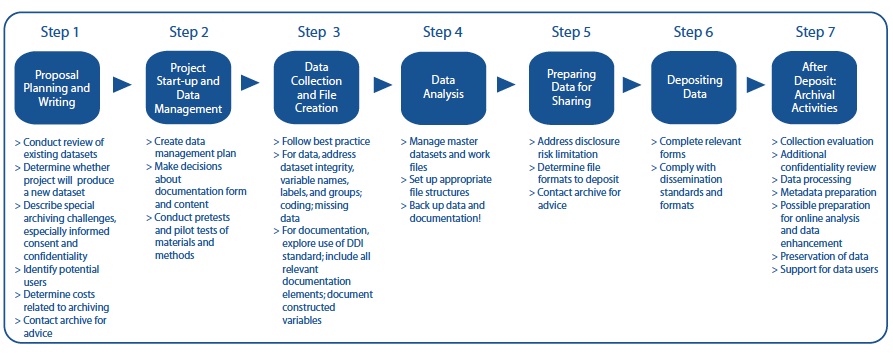


**35. Peter Fox Full Life Cycle of Data**



Presented by Peter Fox, Rennsselaer Polytechnic Institute at USGS CDI August 13, 2010 Denver, CO.

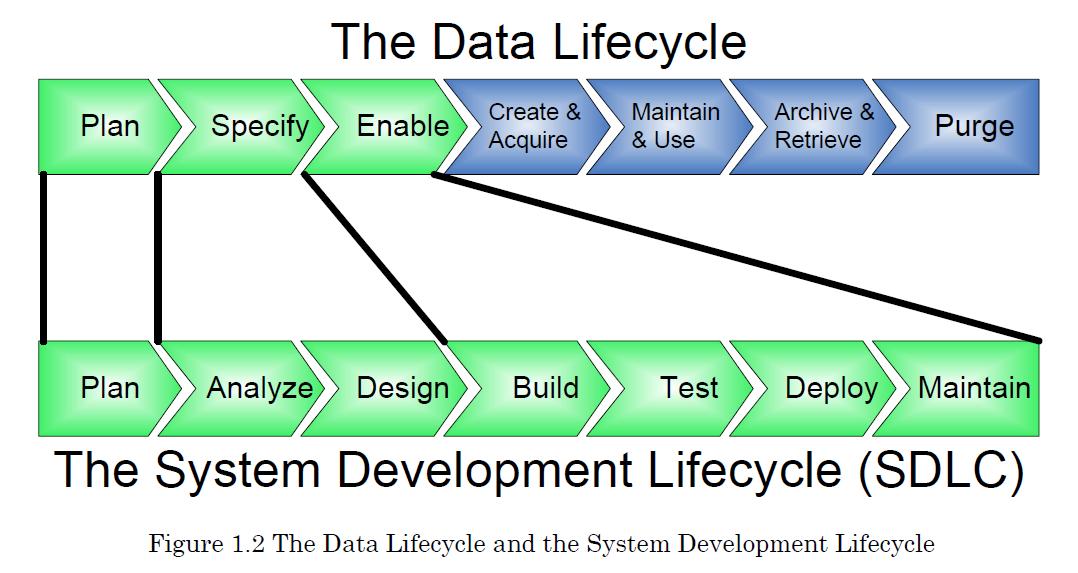
**36. Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle**



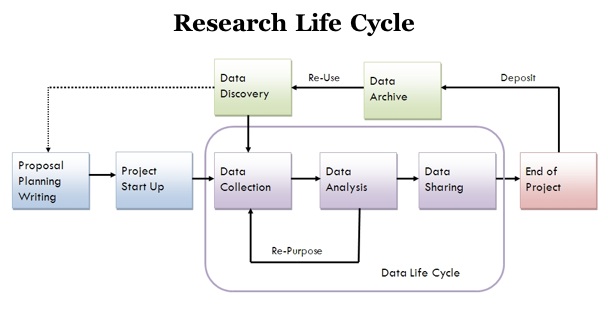
2009 - 4th edition from Inter-University Consortium for Political and Social Research (ICPSR)

Published by: ICPSR Institute for Social Research University of Michigan P.O. Box 1248 Ann Arbor, MI 48106.

**37. Steve Tessler Data and System Lifecycle Models**



**38. University of Virginia Library**



Obtained 8-12-2011 from University of Virginia Library Scientific Data Consulting site URL http://www2.lib.virginia.edu/brown/data/

**39. Mario Valle**

**Scientific Data Management**

Someone said: “We are drowning in data, but starving of information”. And this is particularly true for scientific data. This happens also for business data, but here they had more time to learn. They implemented data architectures, created data warehouse and used data mining to extract information from their data. So why don't study and implement something similar for scientific data? The solution can be to setup a Scientific Data Management architecture.

Scientists normally limit the meaning of Data Management to the mere physical data storage and access layer. But the scope of Scientific Data Management is much boarder: it is about **meaning** and **content**.

Below I listed common problem and opportunities in scientific data access. Then I collected what are considered the parts of a Data Management solution. A list of references and examples of data access and scientific data collections follow.

The paper ends with more implementation oriented issues: a survey of some scientific data formats, planning for a possible implementation and a survey of the supporting technologies available.

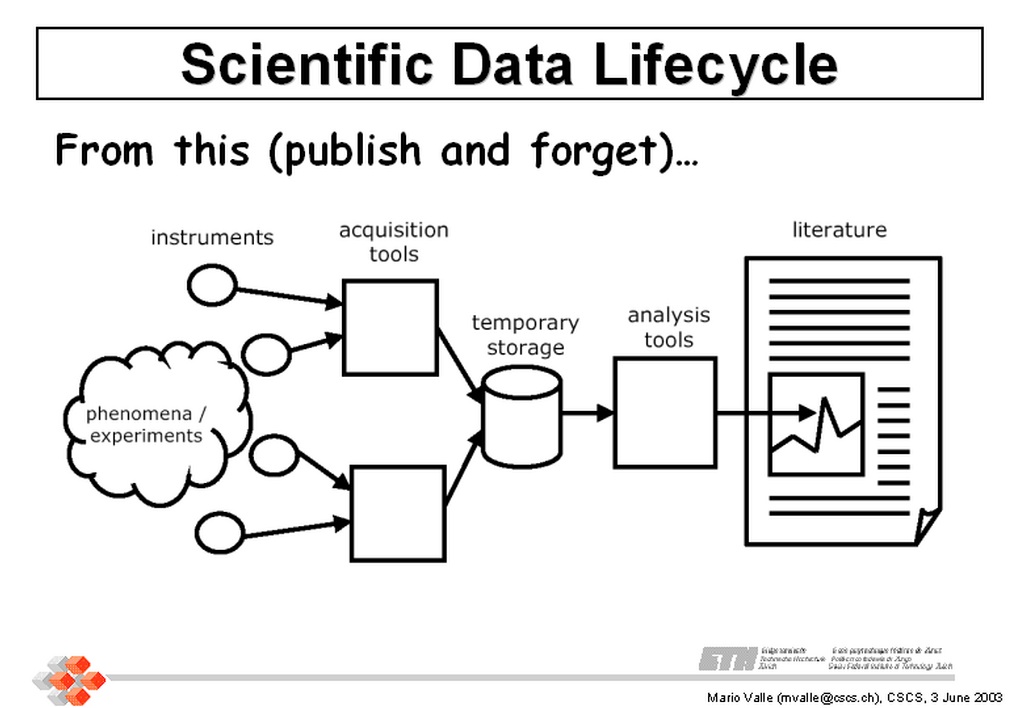
Most of this paper notes and information have been collected and studied for one specific [project](http://lmhwww.epfl.ch/Publications/Theses/Perrig/PerrigPhD.html). But really the ideas collected are generally applicable to the kind of scientific projects that uses the [CSCS](http://www.cscs.ch/) computational and visualization services.

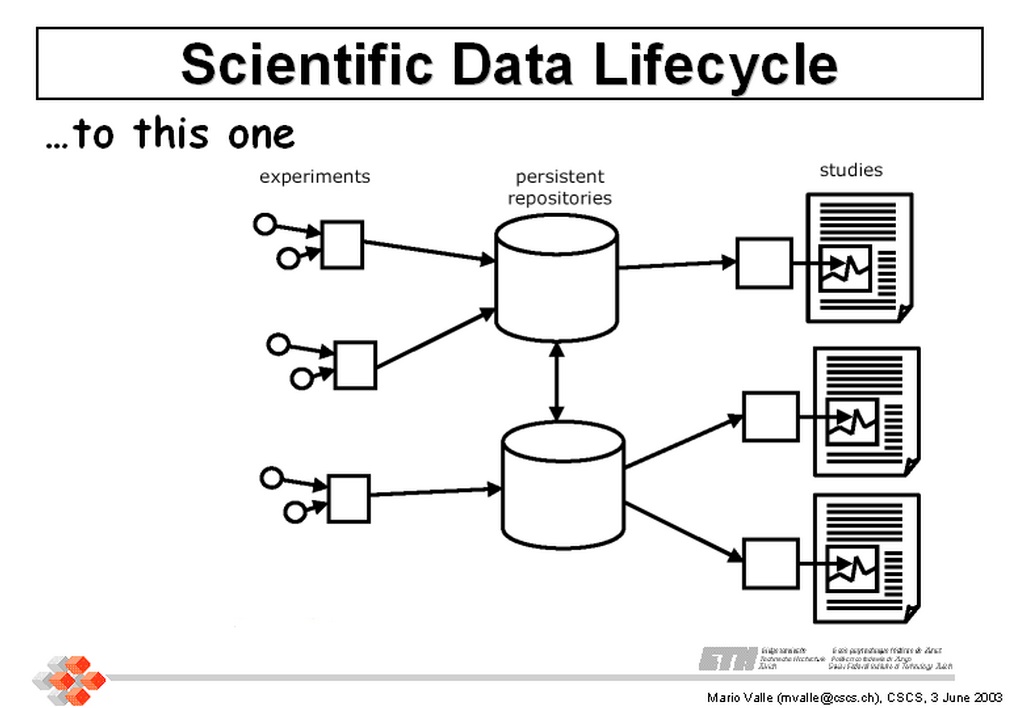
**Problems and opportunities**

Problems that can be found in current scientific projects are for example:

* Limited file and directory naming schemes. Some project data repositories are simply big flat directories.
* Scientists retrieve entire files to ascertain relevance.
* No access to important metadata in scientists' notebooks and heads.
* Un-owned data with dubious content after the end of project or PhD thesis.

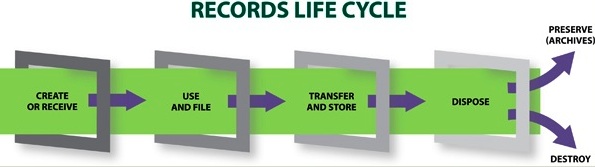
But the increasing of scientific data collections size brings not only problems, but also a lot of opportunities. One of the biggest opportunities is the possibility of reuse existing data for new studies. One example is provided by the various Virtual Observatory initiatives in [Europe](http://www.euro-vo.org/index.html) and [USA](http://us-vo.org/). The idea is summarized below:





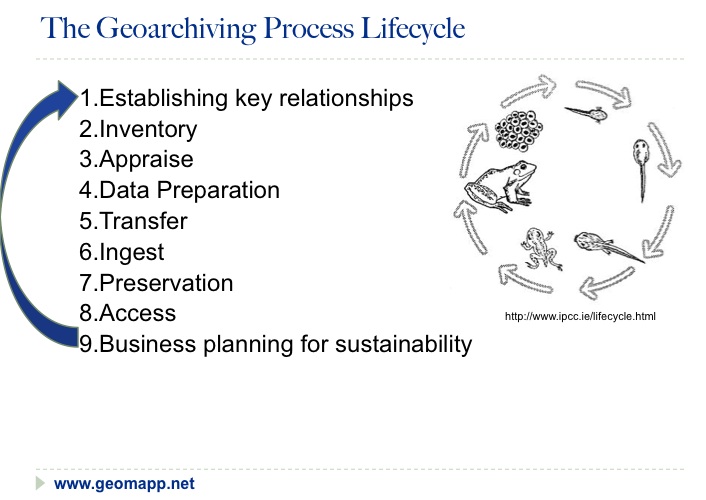
Obtained 8-12-2011 from Mario Valle Web > Scientific Data Management > Introduction URL http://personal.cscs.ch/~mvalle/sdm/scientific-data-management.html

**40. Michigan State University Records Life Cycle Model**



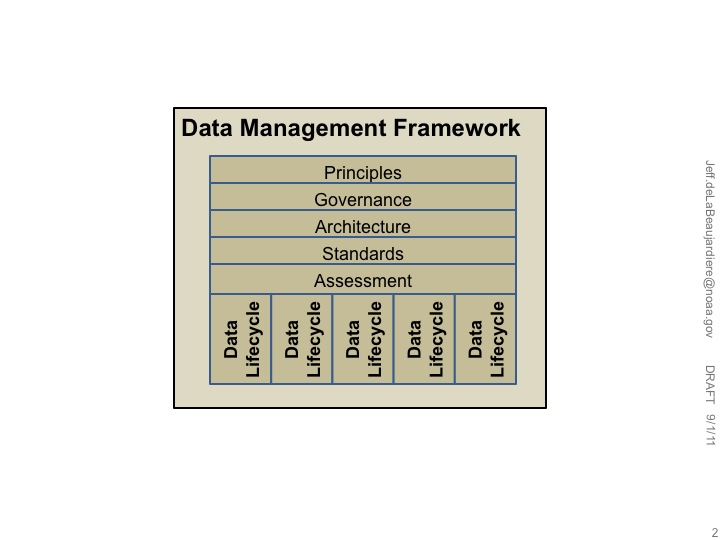
Obtained 8-30-2011 from Michigan State University Archives and Historical Records URL http://archives.msu.edu/records/

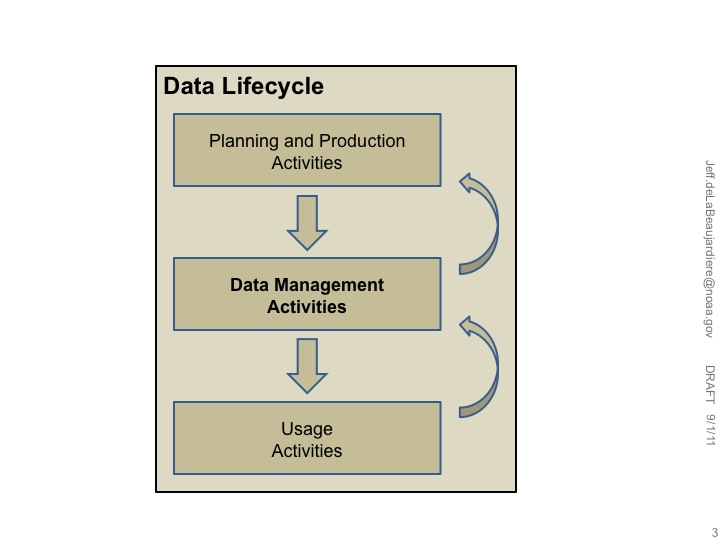
**41. GeoMAPP Geoarchiving Process Lifecycle**

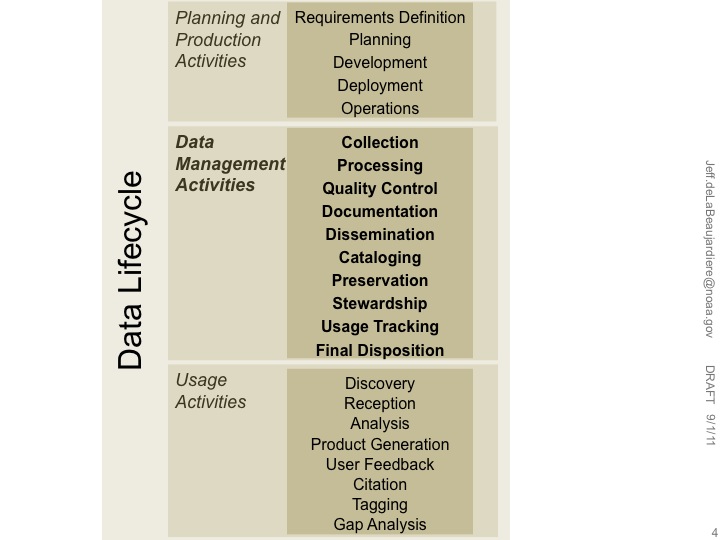


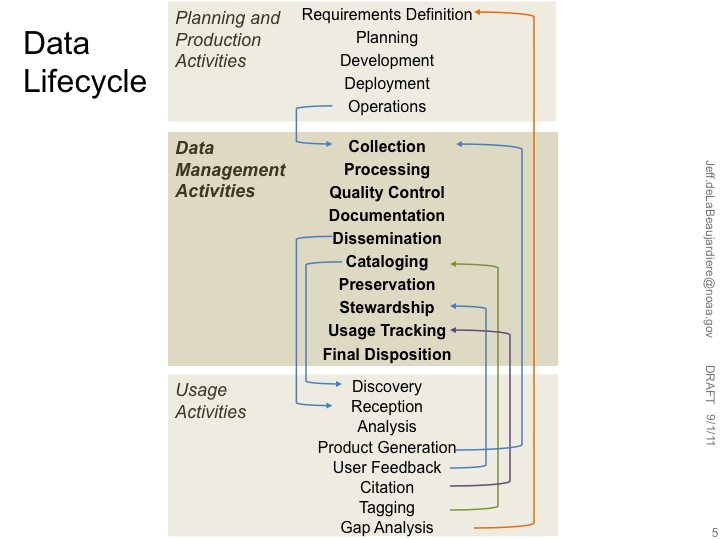
As presented by Steve Morris, North Carolina State University on behalf of GeoMAPP at the Society of American Archivists Annual Conference, August 26, 2011 Chicago, IL.

**42. Jeff de La Beaujardière's proposed Data Management Framework**

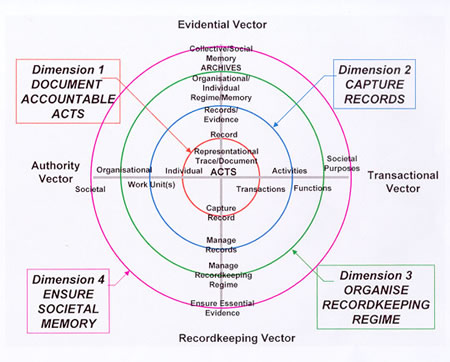








**43. Sarah Demb Continuumm Model**



Obtained 9-8-2011 through correspondence with Sarah Demb.

URL http://john.curtin.edu.au/society/graphics/continuumm.jpg

**44. USGS Data Life Cycle Model Elements**

(Still evolving as of 12 September 2011)