Research Lifecycles

Is today's Data Science movement the new science that John Tukey called for in 1961?

“For a long time I have thought I was a statistician, interested in inferences from the particular to the general. But as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt. ... All in all I have come to feel that my central interest is in data analysis, which I take to include, among other things: procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data”

A new science?

There are diverse views as to what makes a science, but three constituents will be judged essential by most, viz:

(a1) intellectual content,
(a2) organization in an understandable form,
(a3) reliance upon the test of experience as the ultimate standard of validity.

By these tests mathematics is not a science, since its ultimate standard of validity is an agreed-upon sort of logical consistency and provability.

As I see it, data analysis passes all three tests, and I would regard it as a science, one defined by a ubiquitous problem rather than by a concrete subject. Data analysis and the parts of statistics which adhere to it, must then take on the characteristics of a science rather than those of mathematics, ...

These points are meant to be taken seriously.

The four driving forces of the new science

Four major influences act on data analysis today:

1. Formal theories of statistics,

2. Accelerating developments in computers and display devices,

3. The challenge, in many fields, of more and ever larger bodies of data,

4. The emphasis on quantification in an ever wider variety of disciplines.

SOCIAL SCIENCE

Computational Social Science

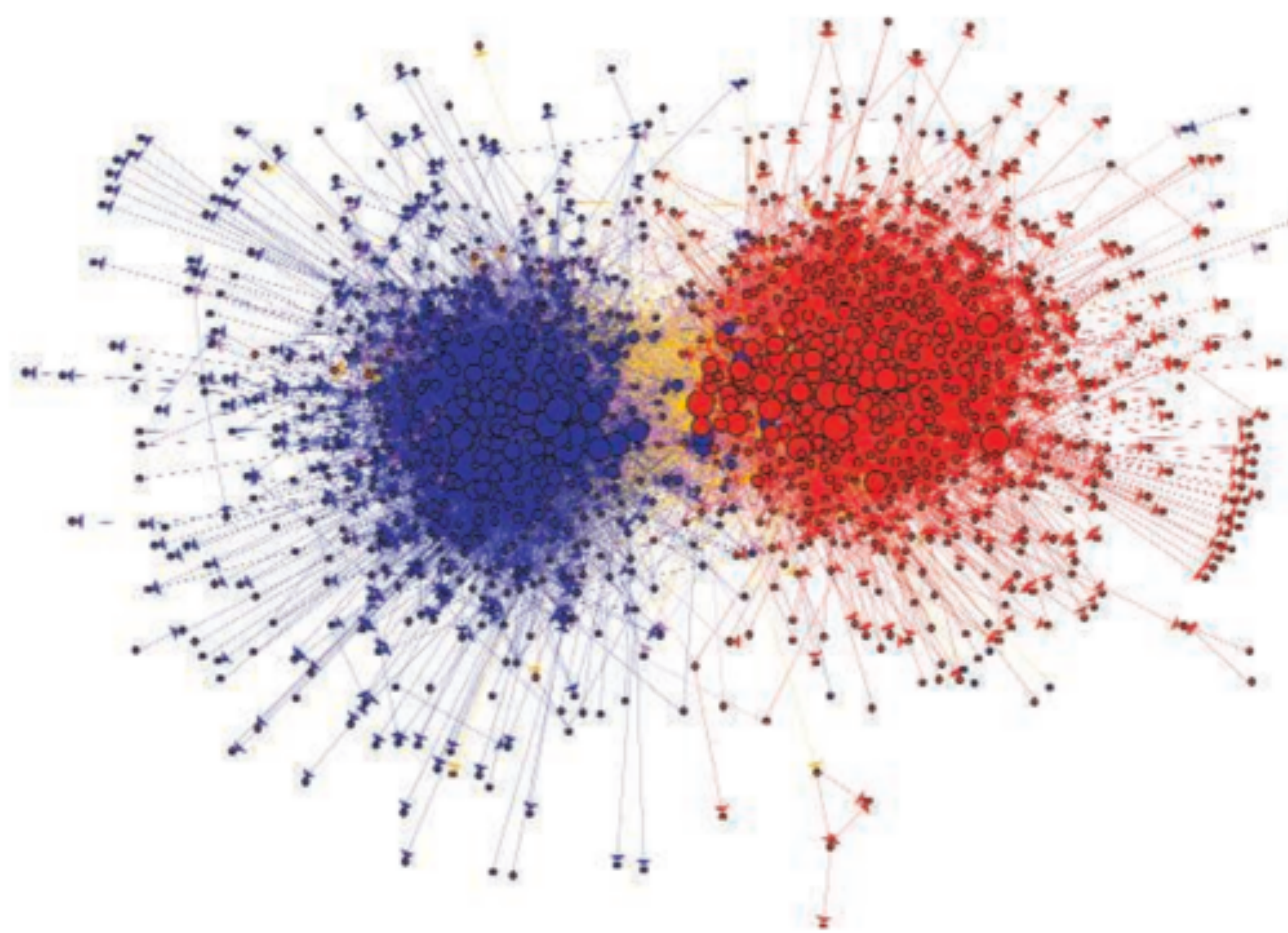
A field is emerging that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors.

David Lazer,[1] Alex Pentland,[2] Lada Adamic,[3] Sinan Aral,[2,4] Albert-László Barabási,[5] Devon Brewer,[6] Nicholas Christakis,[1] Noshir Contractor,[7] James Fowler,[8] Myron Gutmann,[3] Tony Jebara,[9] Gary King,[1] Michael Macy,[10] Deb Roy,[2] Marshall Van Alstyne[2,11]

We live life in the network. We check our e-mails regularly, make mobile phone calls from almost any location, swipe transit cards to use public transportation, and make purchases with credit cards. Our movements in public places may be captured by video cameras, and our medical records stored as digital files. We may post blog entries accessible to anyone, or maintain friendships through online social networks. Each of these transactions leaves digital traces that can be compiled into comprehensive pictures of both individual and group behavior, with the potential to transform our understanding of our lives, organizations, and societies.

The capacity to collect and analyze massive amounts of data has transformed such fields as biology and physics. But the emergence of a data-driven "computational social science" has been much slower. Leading journals in economics, sociology, and political science show little evidence of this field. But computational social science is occurring—in Internet companies such as Google and Yahoo, and in government agencies such as the U.S. National Security Agency. Computational social science could become the exclusive domain of private companies and government agencies. Alternatively, there might emerge a privileged set of academic researchers presiding over private data from which they produce papers that cannot be critiqued or replicated. Neither scenario will serve the long-term public interest of accumulating, verifying, and disseminating knowledge.

What value might a computational social science—based in an open academic environment—offer society, by enhancing understanding of individuals and collectives? What are the



Data from the blogosphere. Shown is a link structure within a community of political blogs (from 2004), where red nodes indicate conservative blogs, and blue liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it. [Reproduced from (*8*) with permission from the Association for Computing Machinery]

[1]Harvard University, Cambridge, MA, USA. [2]Massachusetts Institute of Technology, Cambridge, MA, USA. [3]University of Michigan, Ann Arbor, MI, USA. [4]New York University, New York, NY, USA. [5]Northeastern University, Boston, MA, USA. [6]Interdisciplinary Scientific Research, Seattle, WA, USA. [7]Northwestern University, Evanston, IL, USA. [8]University of California–San Diego, La Jolla, CA, USA. [9]Columbia University, New York, NY, USA [10]Cornell University, Ithaca, NY, USA. [11]Boston University, Boston, MA, USA. E-mail: david_lazer@harvard.edu. Complete affiliations are listed in the supporting online material.

Published by AAAS

Tukey and Wilk 1969

... data analysis is a very difficult field. It must adapt itself to what people can and need to do with data. In the sense that biology is more complex than physics, and the behavioral sciences are more complex than either, it is likely that the general problems of data analysis are more complex than those of all three. It is too much to ask for close and effective guidance for data analysis from any highly formalized structure, either now or in the near future.

Data analysis can gain much from formal statistics, but only if the connection is kept adequately loose.

Data Science Pipeline?

1. Data collection, cleaning, preprocessing (experimental design)
2. Data exploration
3. Workflow tracking and provenance tools
4. Inference, modeling, prediction.. Computing with data
5. Data visualization and presentation
6. Data and code for sharing: metadata, documentation, workflows, publications, discoverability..
7. Data, code, ideas re-use..

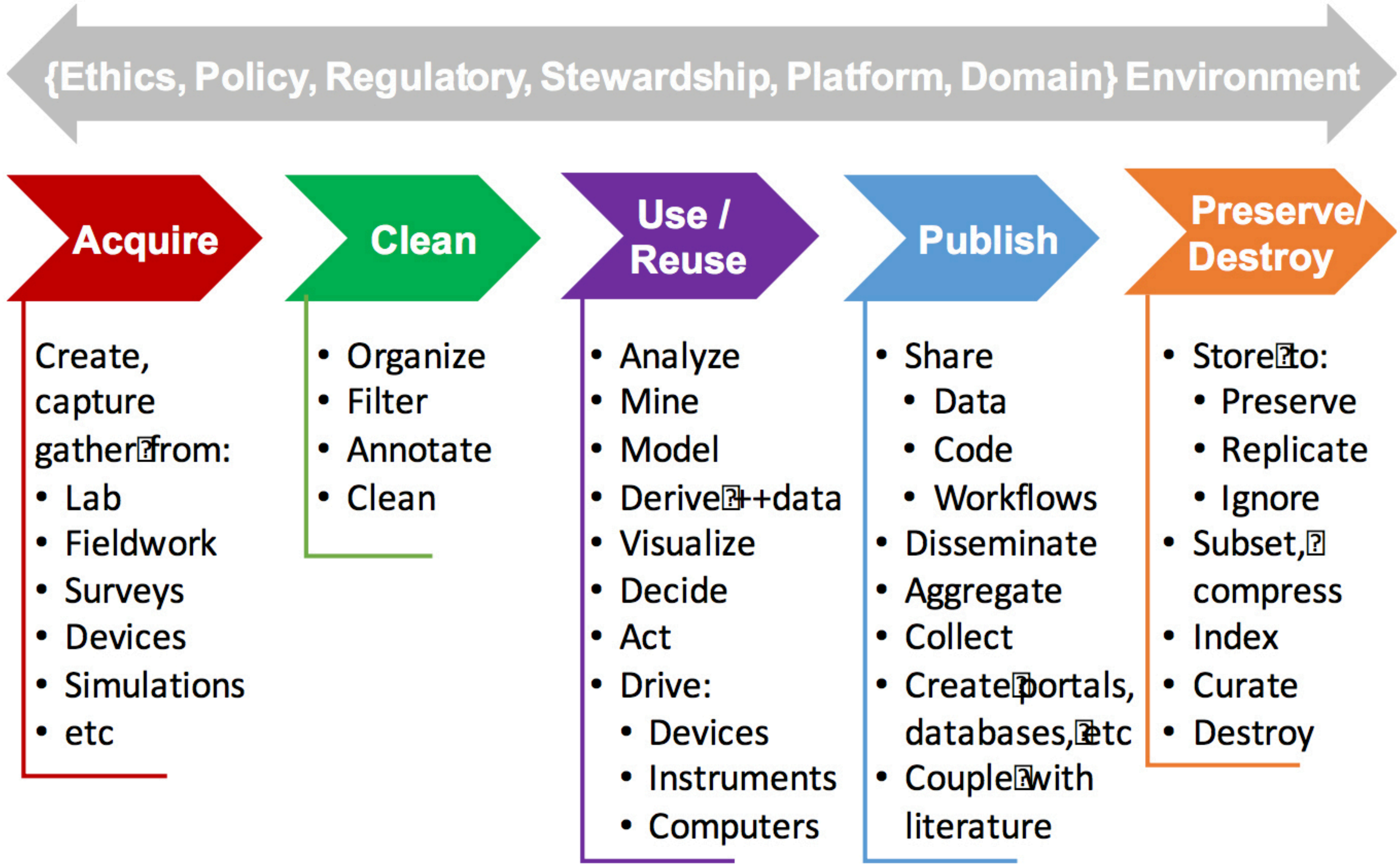


FIGURE 1: The Data Life Cycle and Surrounding Data Ecosystem

The *Data Life Cycle* is critical to understanding the opportunities and challenges of making the most of digital data. **Figure 1** shows a simplified cartoon with essential components of the data life cycle. Data is *acquired* from some source (measured, observed, generated), *cleaned* and edited to remove the outliers inevitable in real-world measurement scenarios and render it suitable for subsequent analysis; *used* (or reused) via some analysis leading to insight, action, or decision; *published* or disseminated in some way so the community at large is made aware of the data and its outcome(s); *preserved* (or not) so that others can revisit and reuse this data now or in the future. Surrounding this overall pipeline is a broader *environment* of concerns: *stewardship* to maximize the quality of the data and promote effective use, *ethics* issues that touch on proper or improper actions with these data; *policy* and *regulatory* constraints that impose legal limitations on these data; *platform* and infrastructure issues that affect technically how we can work with data; and *domain* and disciplinary needs specific to the application communities that create, operate, and use the data from these pipelines.

The Future of Data Analysis

From Efron's talk from the early 1980's - 'A Theory of Statistics based on Massive Computation:'

"every meaningful performance goal which had classically been addressed by mathematical assumptions and theoretical derivations, would in the future be addressed using empirical data and heavy computation."

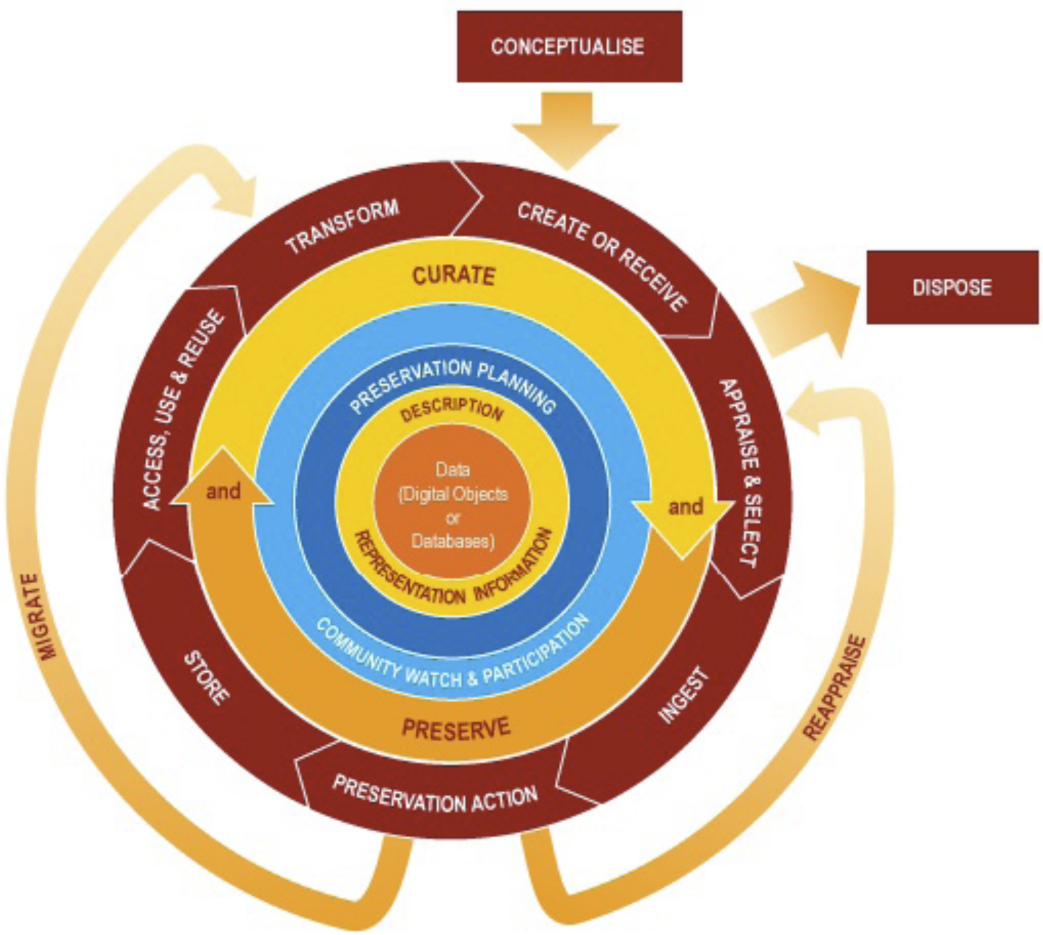
CONTENTS

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.

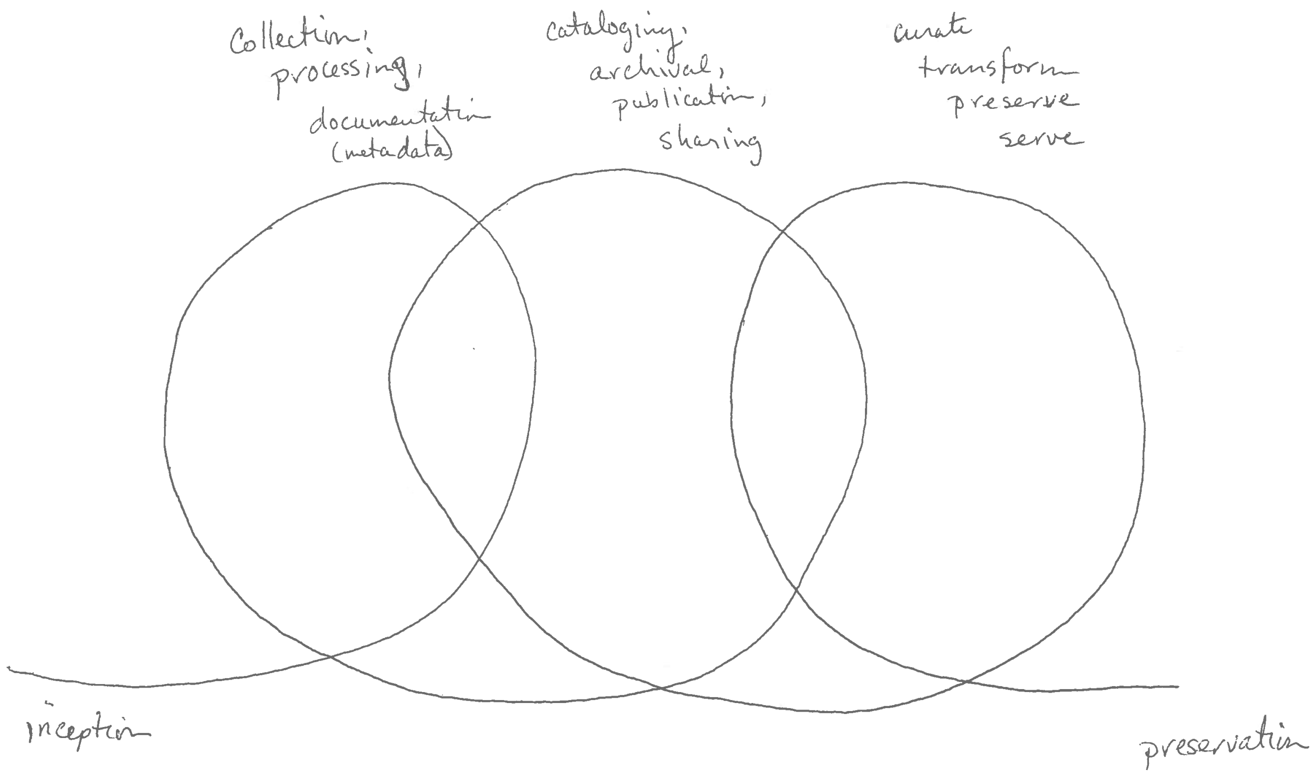
Migrate

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.

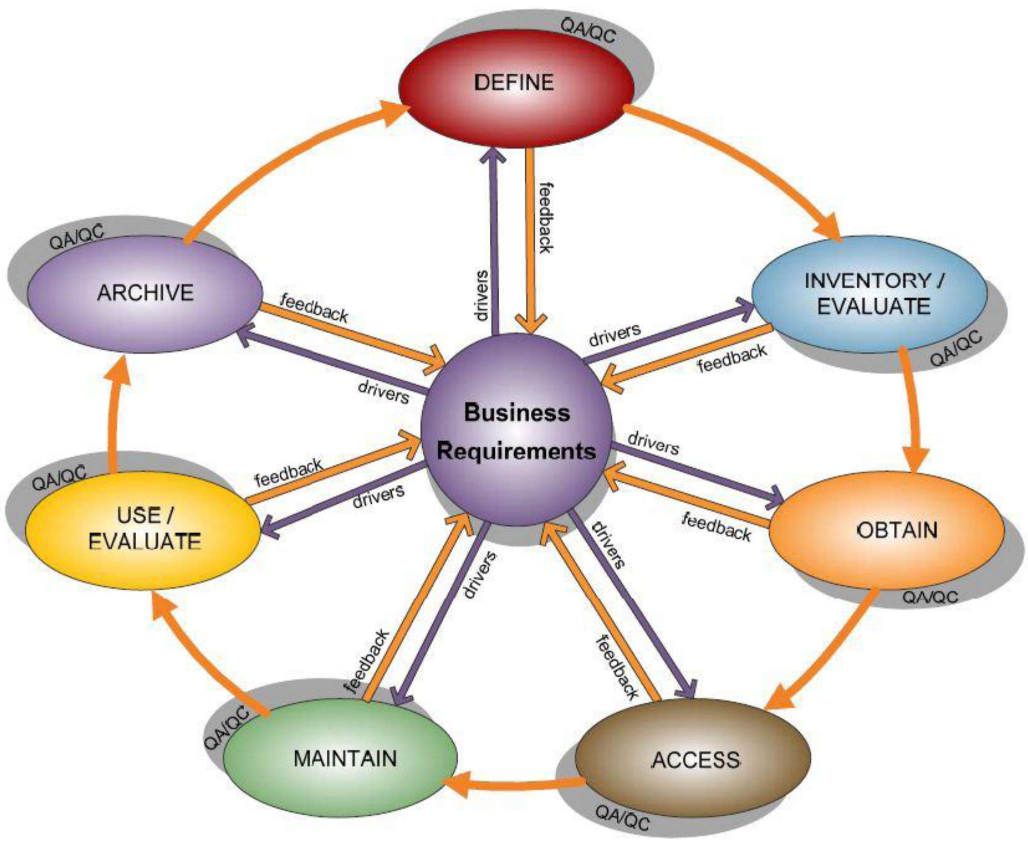


The ideas I tried to capture are:

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

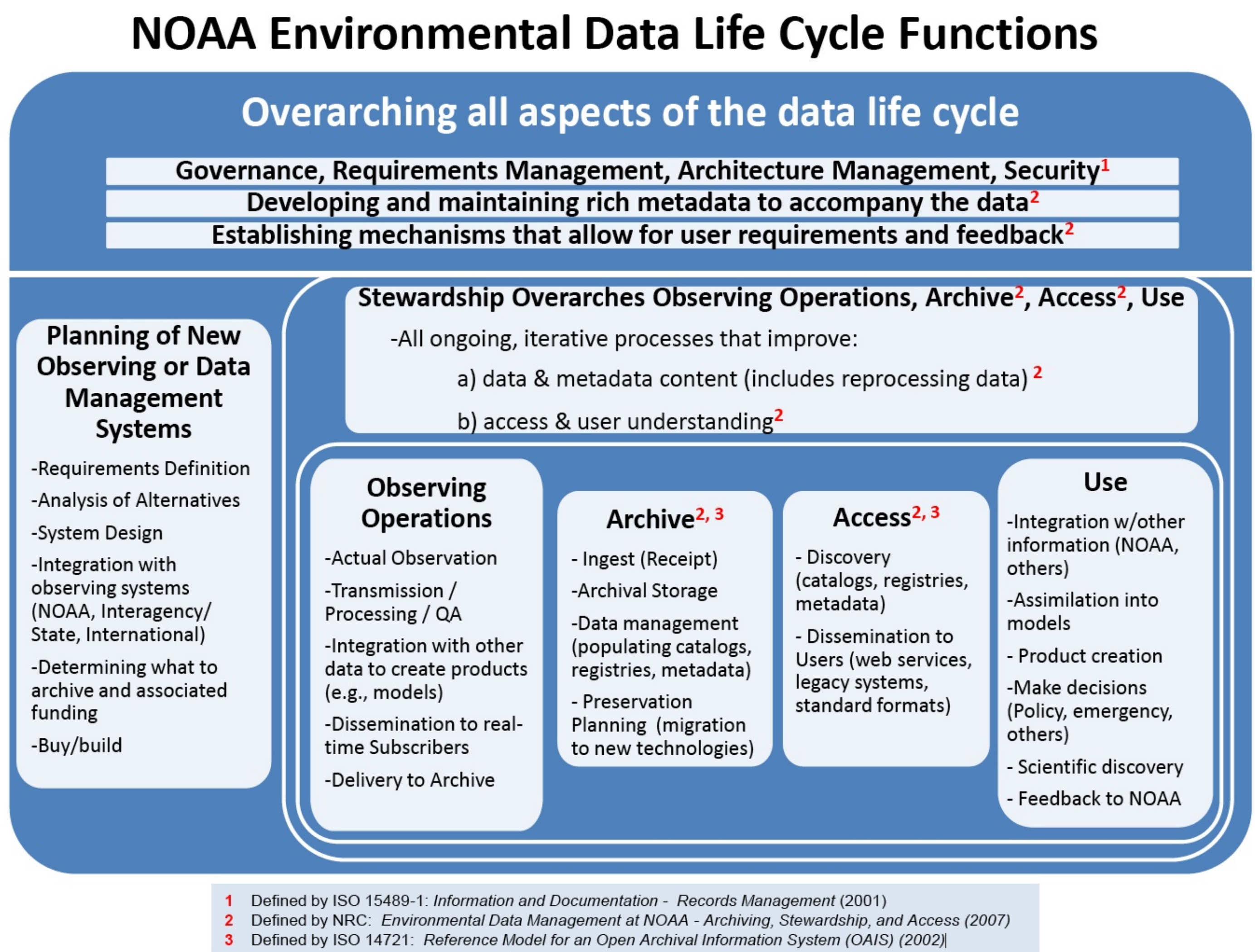
Found at http://www.admin.ox.ac.uk/rdm/

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 Steuerer, 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

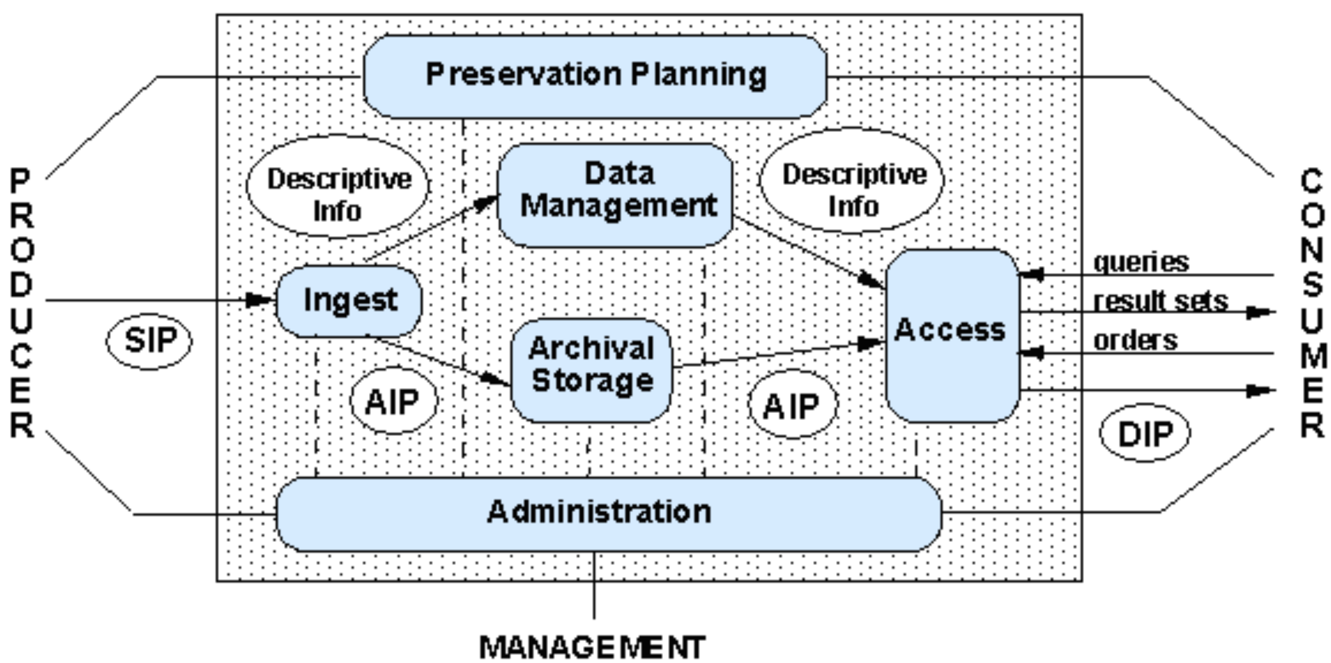
OAIS Framework for Developing a "Community of Practice"

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



Source: Procedures Manual for the Consultative Committee for Space Data Systems (2001)

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

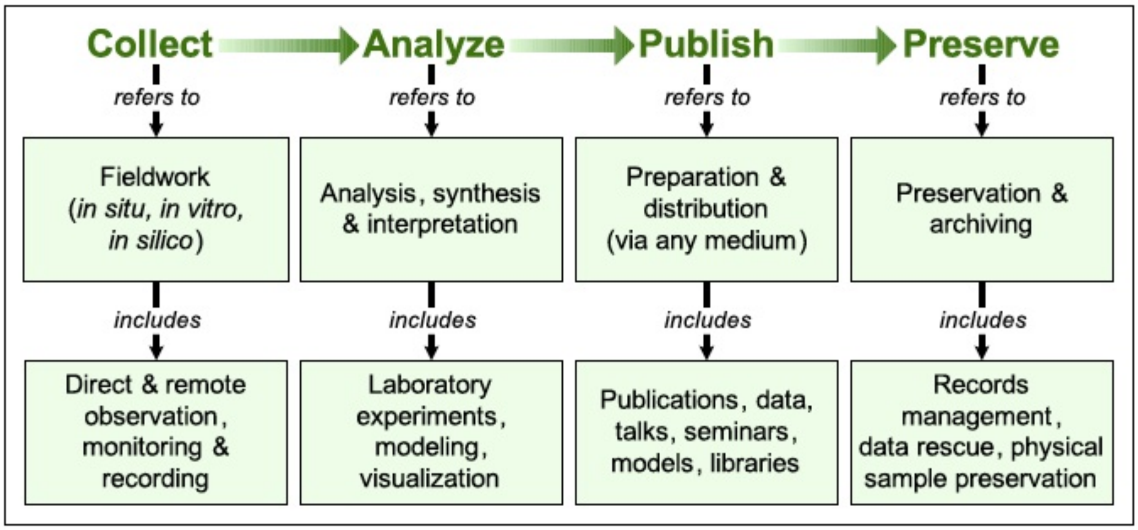
- No integrated “community of practice” for promoting trusted repositories
- CCDSC Red Book and DRAMBORA are not widely known in the scientific, IT, and data management communities
- They do not address design of data systems and data management, e.g. systems planning, funding, system requirements
- No Case Studies and Lessons Learned

7. USGS Scientific Information Management Workshop Vocabulary

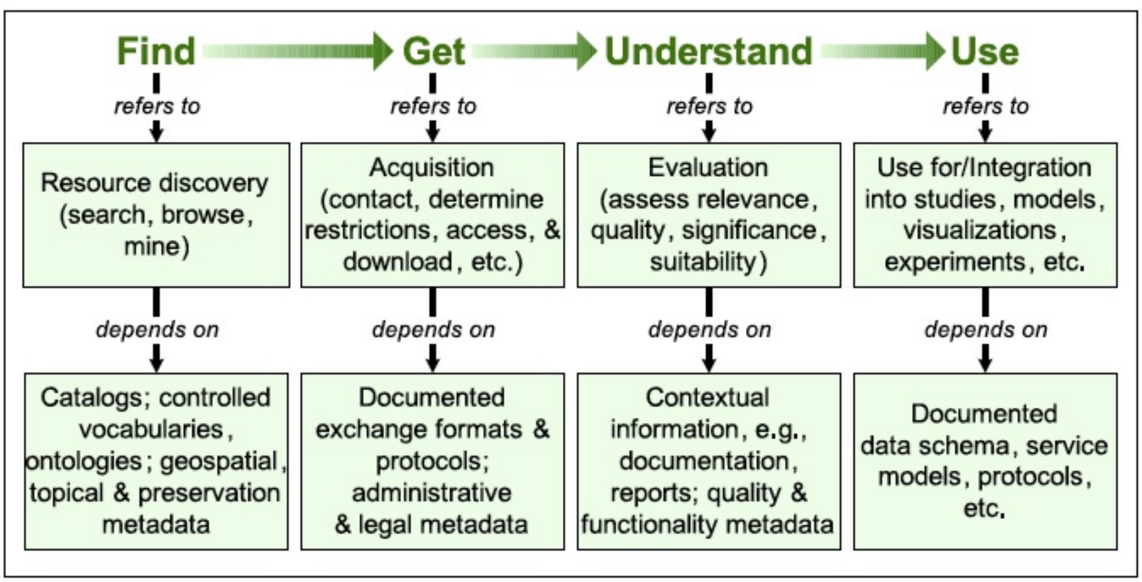
From Tom Gunther and Dave Govoni, USGS, 2006.

A Vocabulary for Scientific Information and Knowledge Management

Producer perspective



Consumer perspective

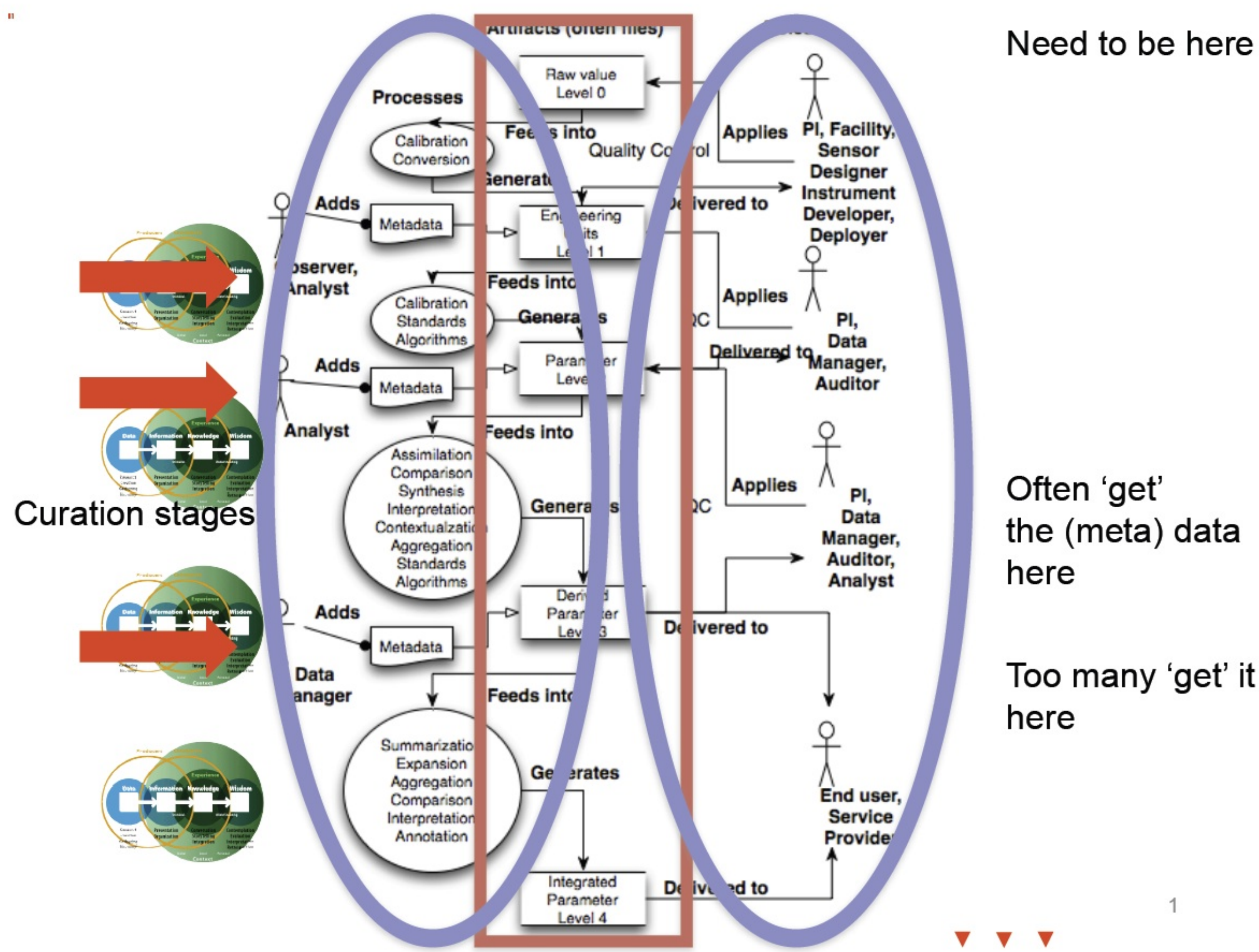


Source: Govoni, D.L. and T.M. Gunther, 2006. Scientific Information Management at the U.S. Geological

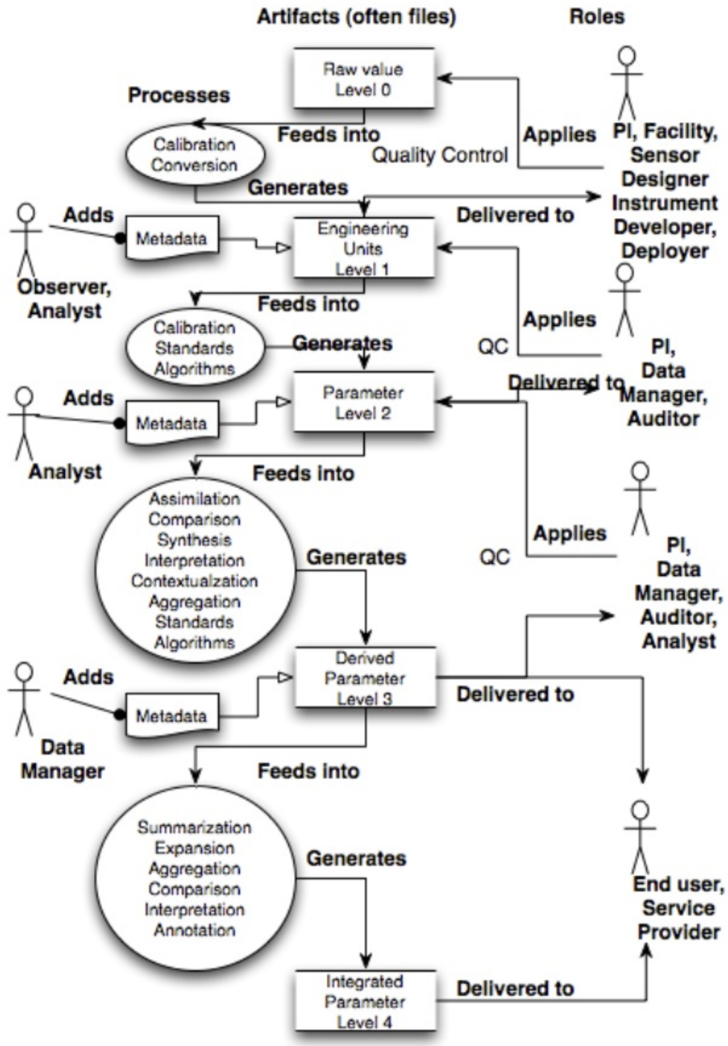
Survey: Issues, Challenges, and a Collaborative Approach to Identifying and Applying Solutions (Abstract). Geoinformatics 2006—Abstracts. Scientific Investigations Report 2006-5201, p. 19-20. U.S. Geological Survey, Reston, Virginia.

8. Peter Fox Lifecycle Diagrams

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



- But back to reality
- Fragmentation
- Disconnection
- Encapsulation
- … all are bad for … **transparency**



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.

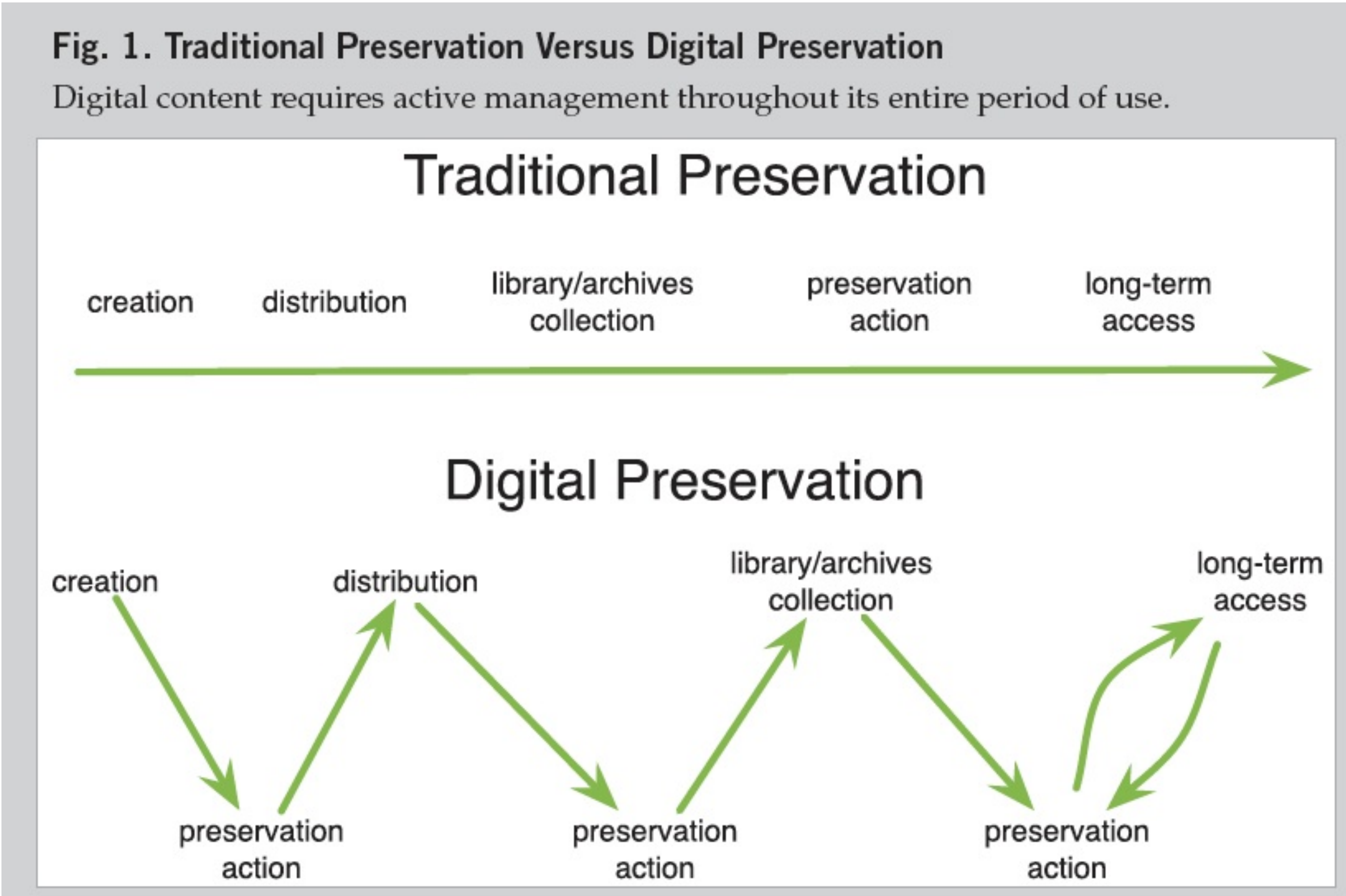
* Data protection - Provide systems, tools, policies, and procedures for protecting legitimate privacy, confidentiality, intellectual property, or other security needs as appropriate to the data type and use.

* 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

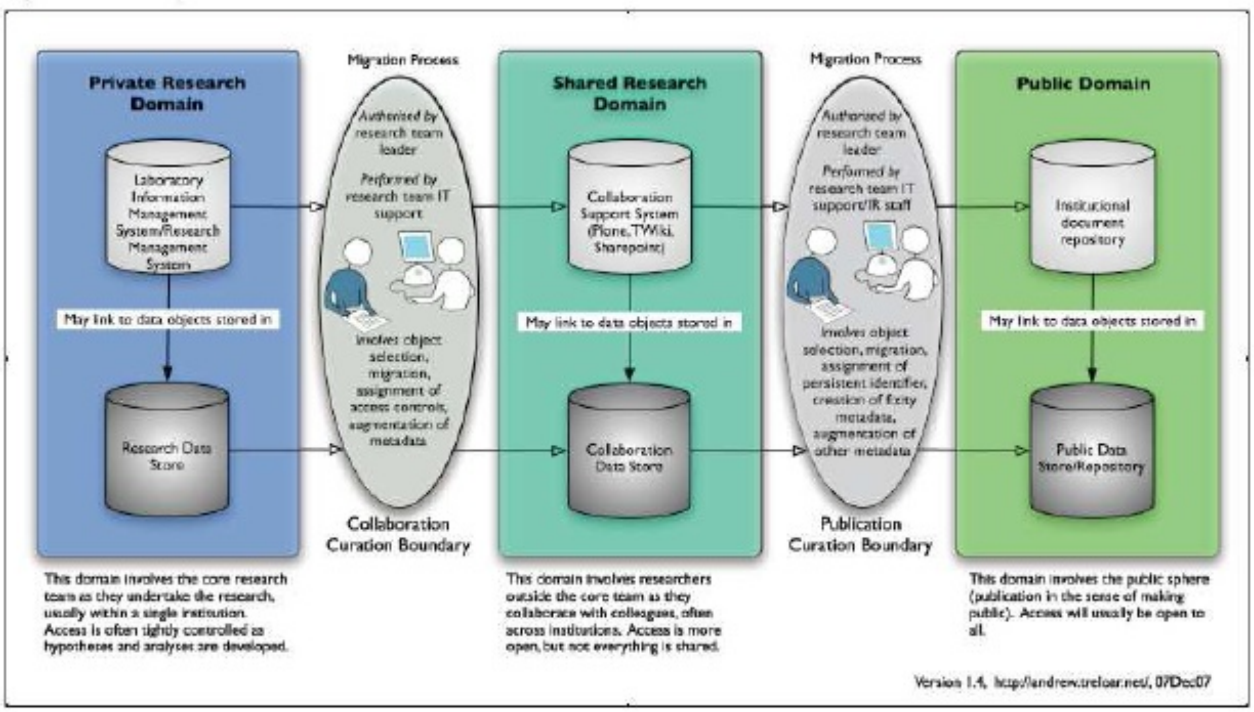
A literature study of researchers' requirements with respect to storage and access to research data.

SURFfoundation
PO Box 2290
NL-3500 GG Utrecht

Author Martin Feijen

February 2011

Figure 1: Domains, Data Stores and Curation Boundaries



Source: Treloar, A. and Harboe-Ree, C. (2008). "Data management and the curation continuum: how the Monash experience is informing repository relationships". *Proceedings of VALA 2008*, Melbourne, February.

12. EPA Project Life Cycle

U.S. Environmental Protection Agency

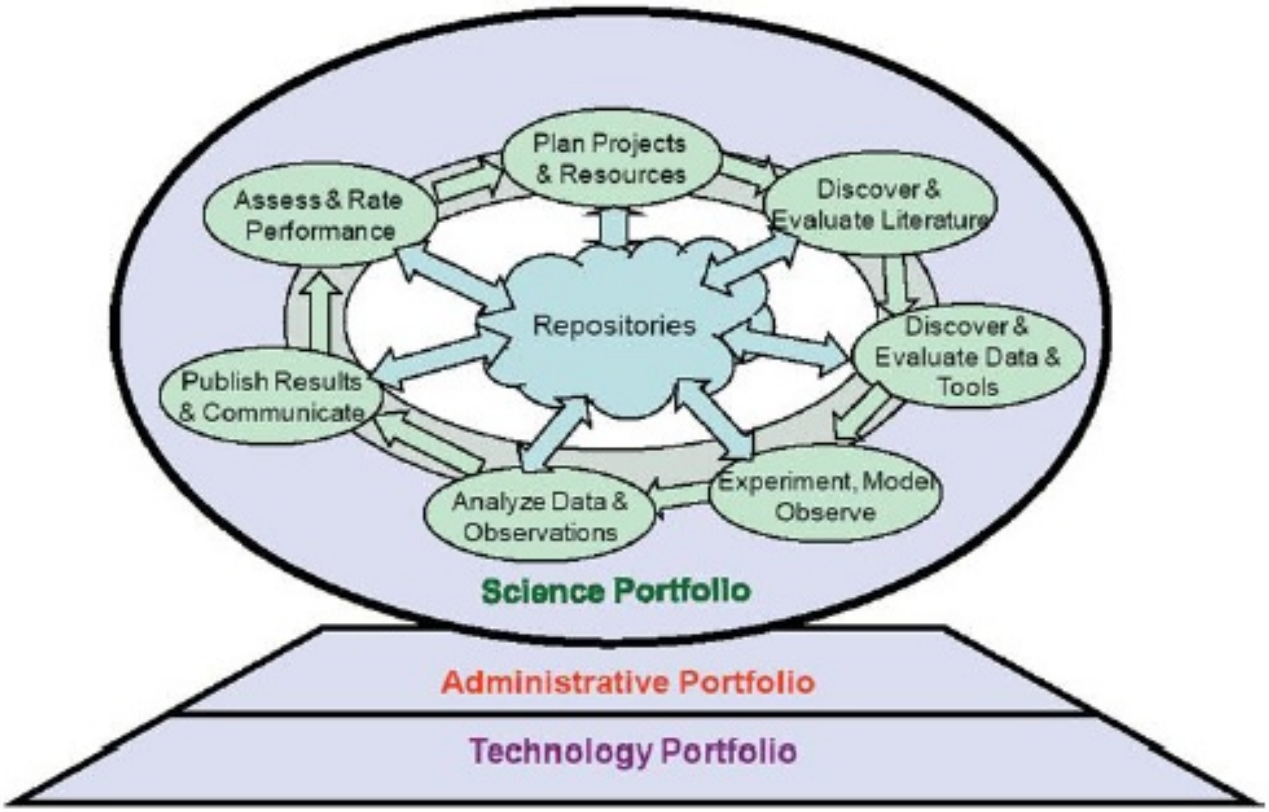


Figure 1.4-1. EPA project lifecycle.

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.]

Figure 1.4-2. IWGDD digital data lifecycle model (NSTC, 2009).

―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.

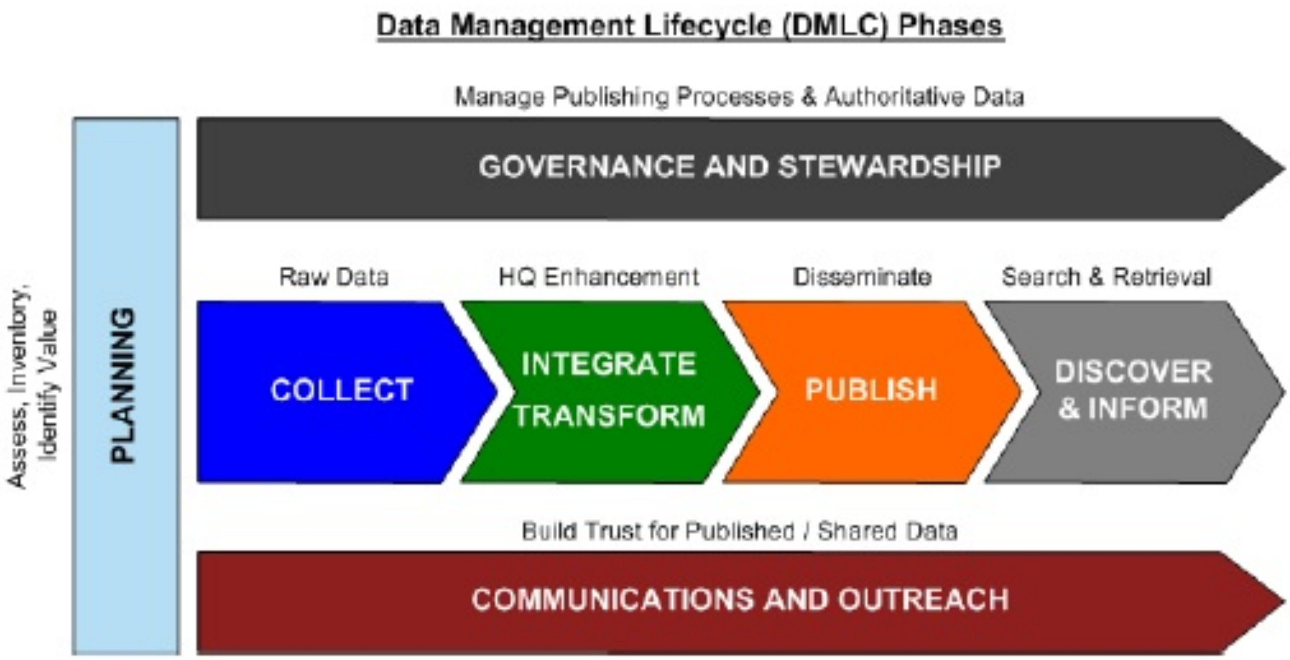


Figure A2. A linear data lifecycle for use in an operational environment (EPA, 2010b).

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. 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. 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. 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. 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. 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. 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)

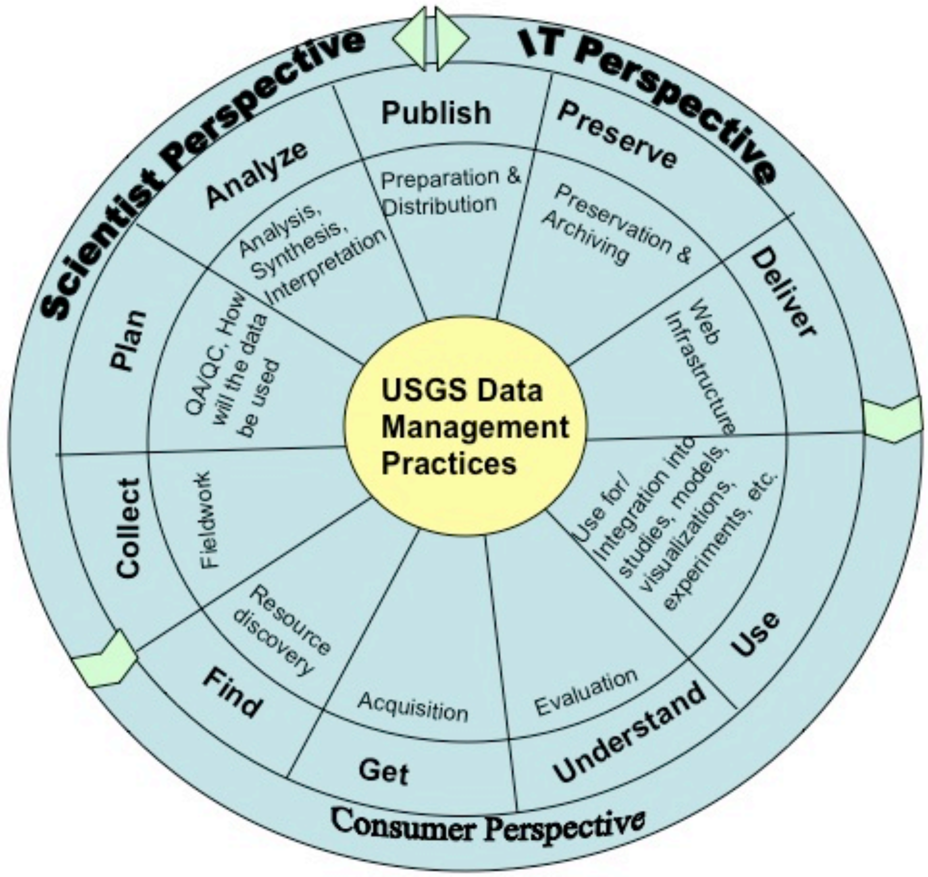
16. Generic Science Data Lifecycle

Figure C2. The 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.

<u>Phase 1</u>
Creation
End User Perspective

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

<u>Phase 2</u>
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.

<u>Phase 2</u>
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.

<u>Phase 2</u>
Dissemination
Participating Libraries Perspective

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

<u>Phase 2</u>
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.

<u>**Phase 3**</u>
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.

<u>**Phase 3**</u>
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.

<u>**Phase 3**</u>
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.

<u>**Phase 3**</u>
Access
End User Perspective

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

<u>Phase 4</u>
Use
Producing Government Agencies Perspective

Provide documentation, software, technical support and user training.

<u>Phase 4</u>
Use
Central Operational Authority Perspective

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

<u>Phase 4</u>
Use
Participating Libraries Perspective

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

<u>Phase 4</u>
Use
End User Perspective

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

<u>Phase 5</u>
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.

<u>Phase 5</u>
Preservation
Central Operational Authority Perspective

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

<u>**Phase 5**</u>
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.

<u>**Phase 5**</u>
Preservation
End User Perspective

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

<u>**Phase 6**</u>
Evaluation
Producing Government Agencies Perspective

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

<u>**Phase 6**</u>
Evaluation
Central Operational Authority Perspective

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

<u>**Phase 6**</u>
Evaluation
Participating Libraries Perspective

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

<u>**Phase 6**</u>
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

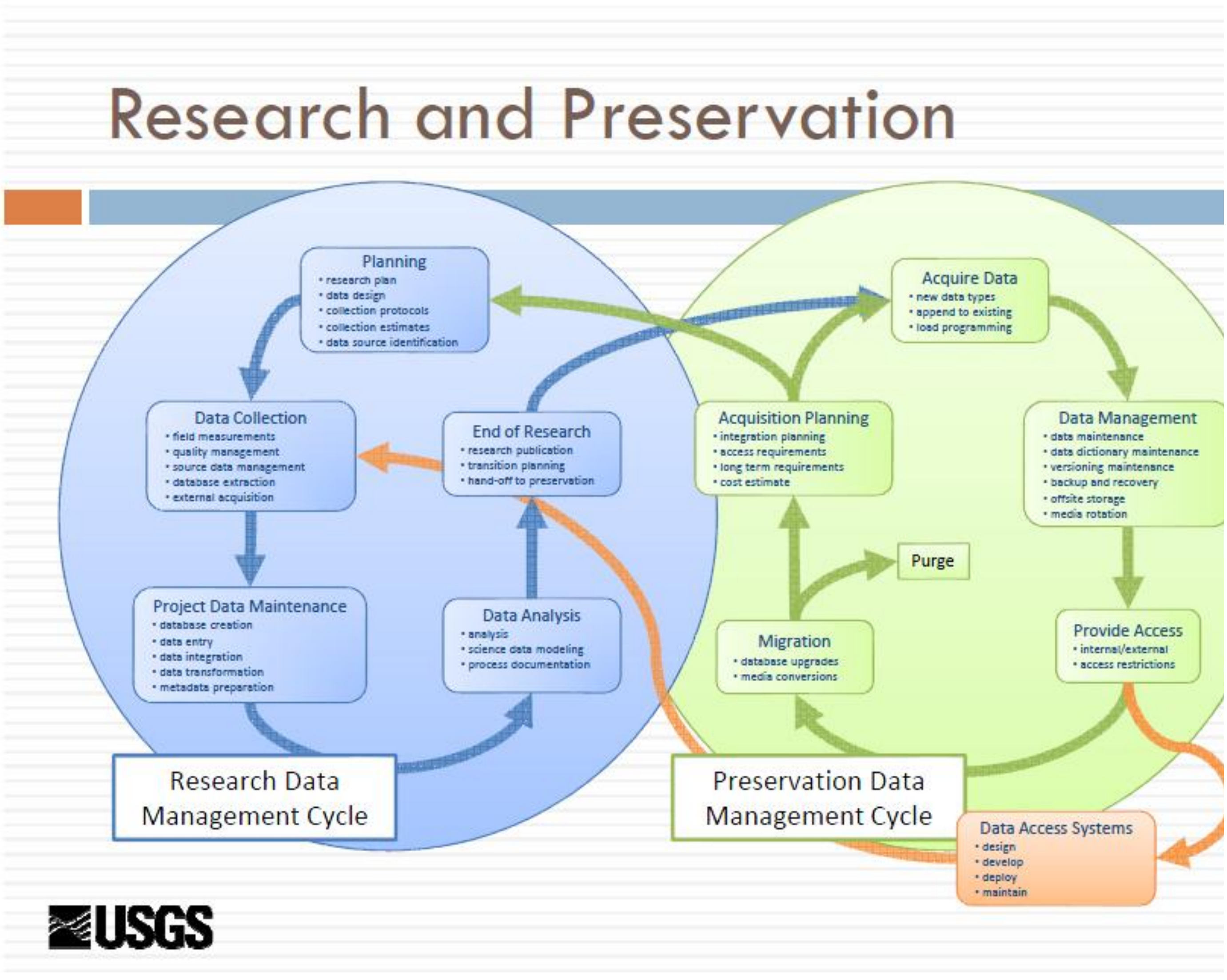
(3) Data Exposure and Delivery

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.

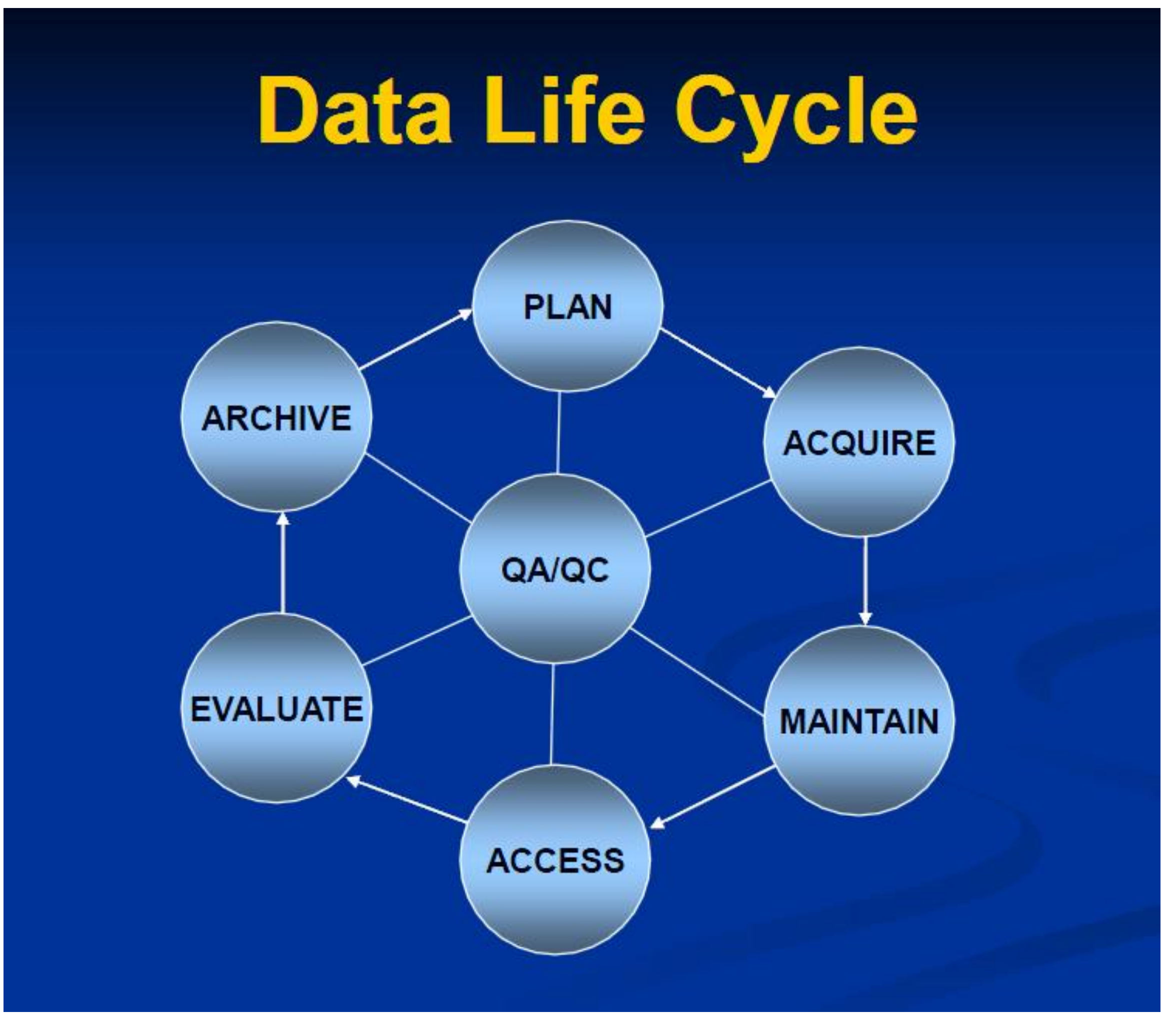
Many opportunities for confusion, conflict, and working at cross-purposes in data management activities can occur when the separate nature of the three DMPf phases

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 <u>each</u> 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:

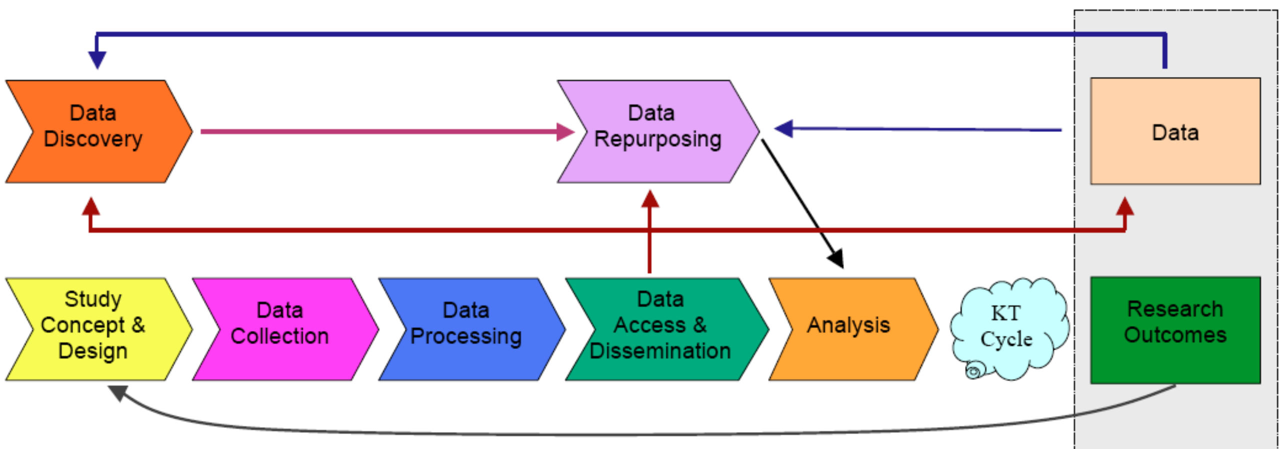
- **Plan**: Decide on the quality level, method for measuring quality and how often the quality should be evaluated.
- **Acquire**: Establish the data acquisition acceptance criteria and the acceptance testing process.
- **Maintain**: Conduct on-going data improvement.
- **Access**: Review and confirm 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

Office of Research Integrity,
U.S. Department of Health and Human Services
Key Components of Data Lifecycle Managemen

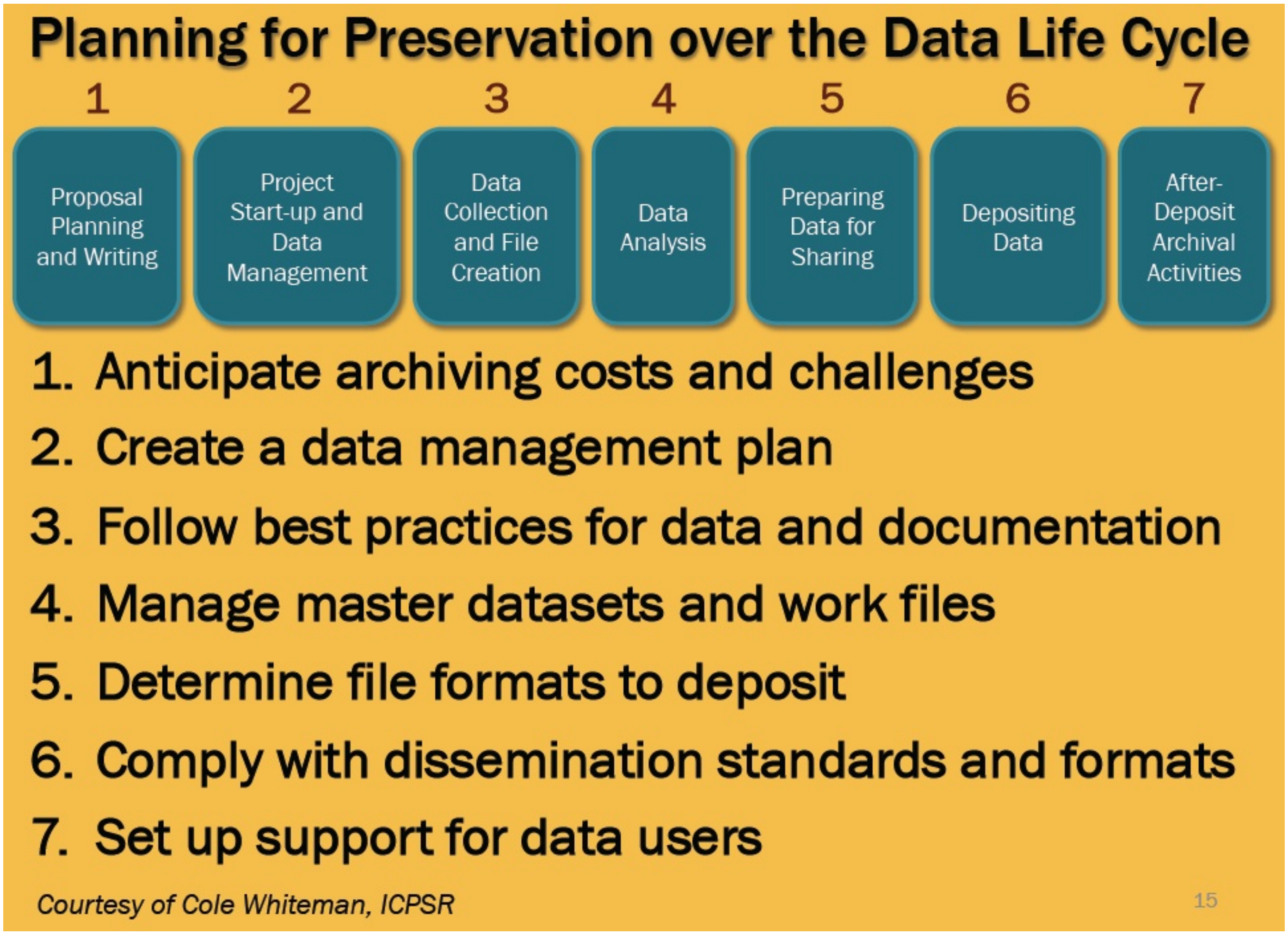
Key Concept	How it Relates to Responsible Conduct of Research
Data Ownership	Concerns who has the legal rights to the data and who retains the data after the project is completed, including the PI's right to transfer their data between institutions
Data Collection	Concerns collecting data in a consistent, systematic manner throughout the project (reliability) and establishing an ongoing system for evaluating and recording changes to the project protocol (validity)
Data Storage	Concerns the amount of data that should be stored - enough so that project results can be reconstructed
Data Protection	Concerns protecting both written and electronic data from physical damage as well as damage to data integrity, including tampering or theft
Data Retention	Concerns how long project data needs to be retained according to various sponsors' and funders' guidelines, and the importance of secure destruction of data
Data Analysis	Concerns how raw data is chosen, evaluated, and interpreted into meaningful and significant conclusions that other researchers and the public can understand and use
Data Sharing	Concerns how project data is disseminated to other researchers and the general public to share important or useful research results; also, when data should not be shared
Data Reporting	Concerns publication of conclusive findings after the project is completed

14

Guidelines for Responsible Data Management in Scientific Research, ori.hhs.gov/education/products/clinicaltools/data.pc

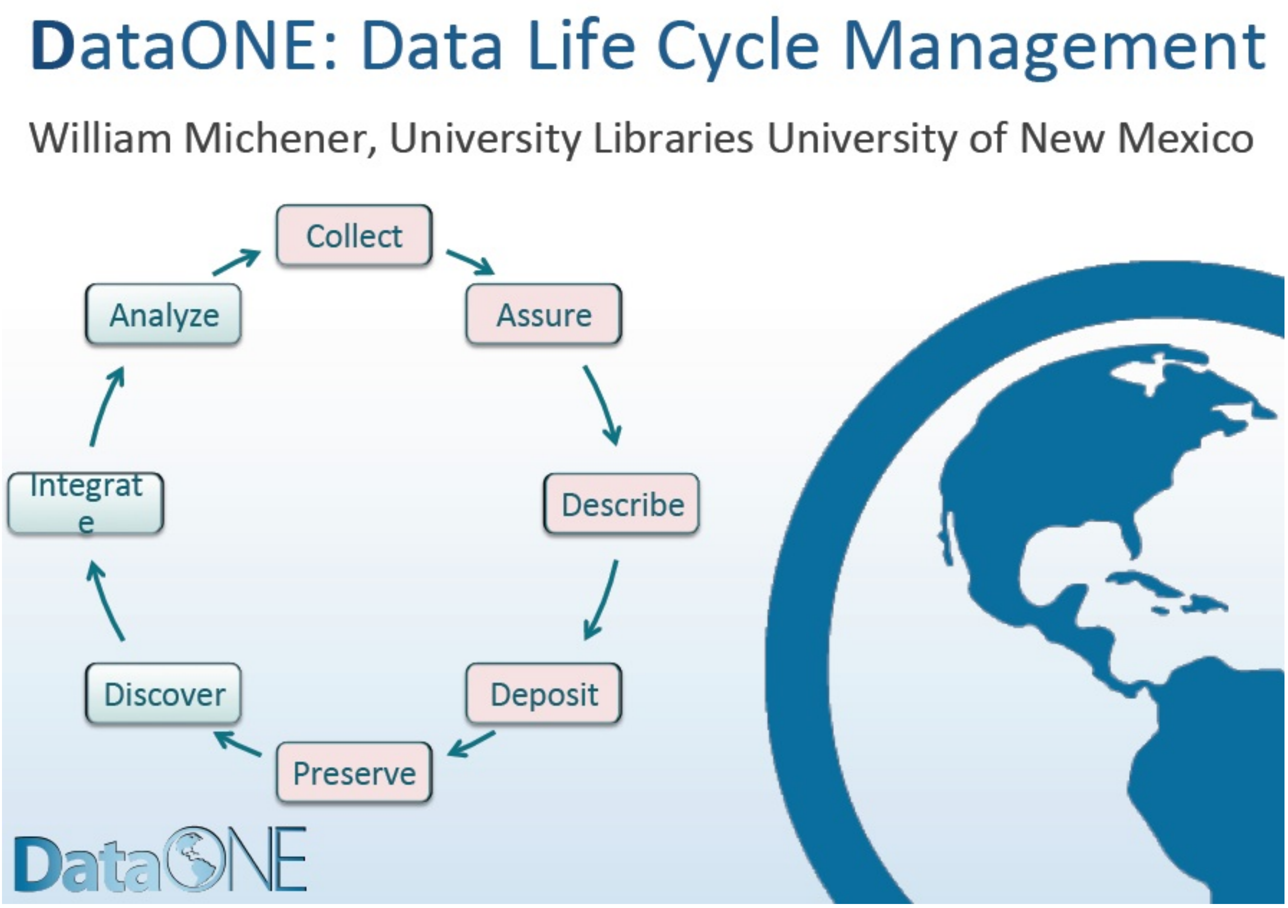
Presented by Jose-Marie Griffiths, Bryant University, at the National Science Foundation Research Data Lifecycle Management Workshop Princeton, NJ July 18-20, 2011.

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.

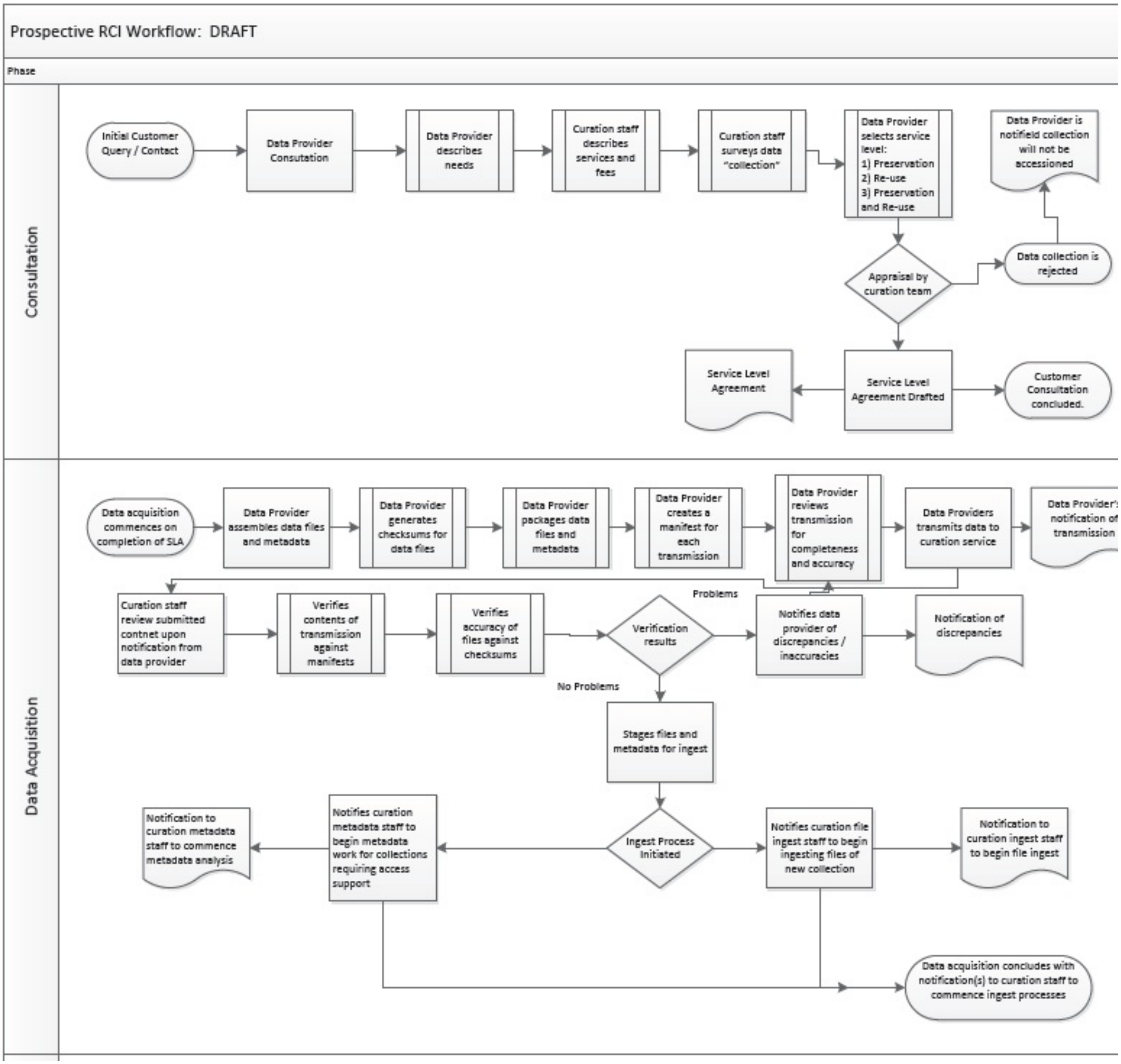
25. IBM Aspects of Lifecycle Management – Research

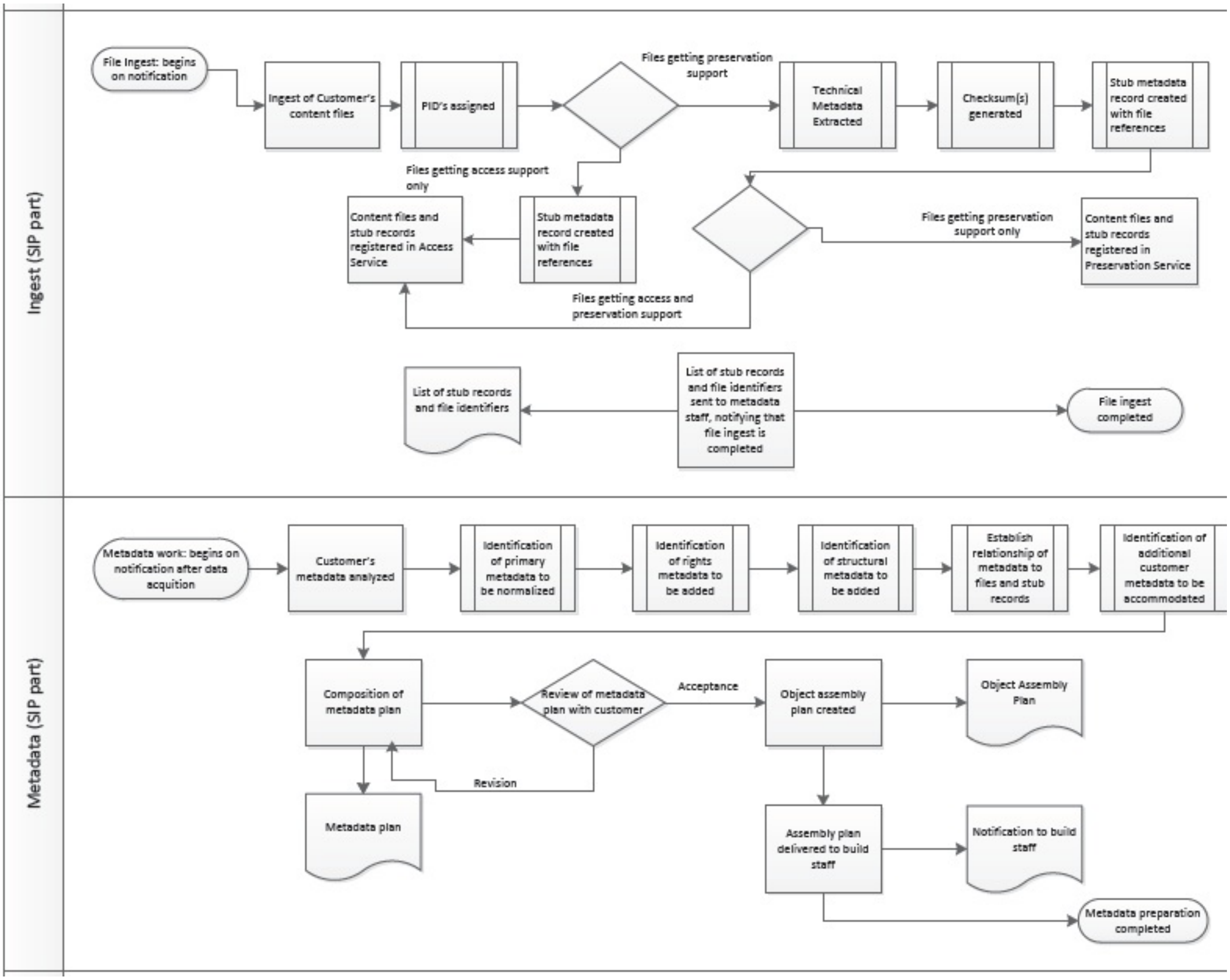
Aspects of Lifecycle Management – Research

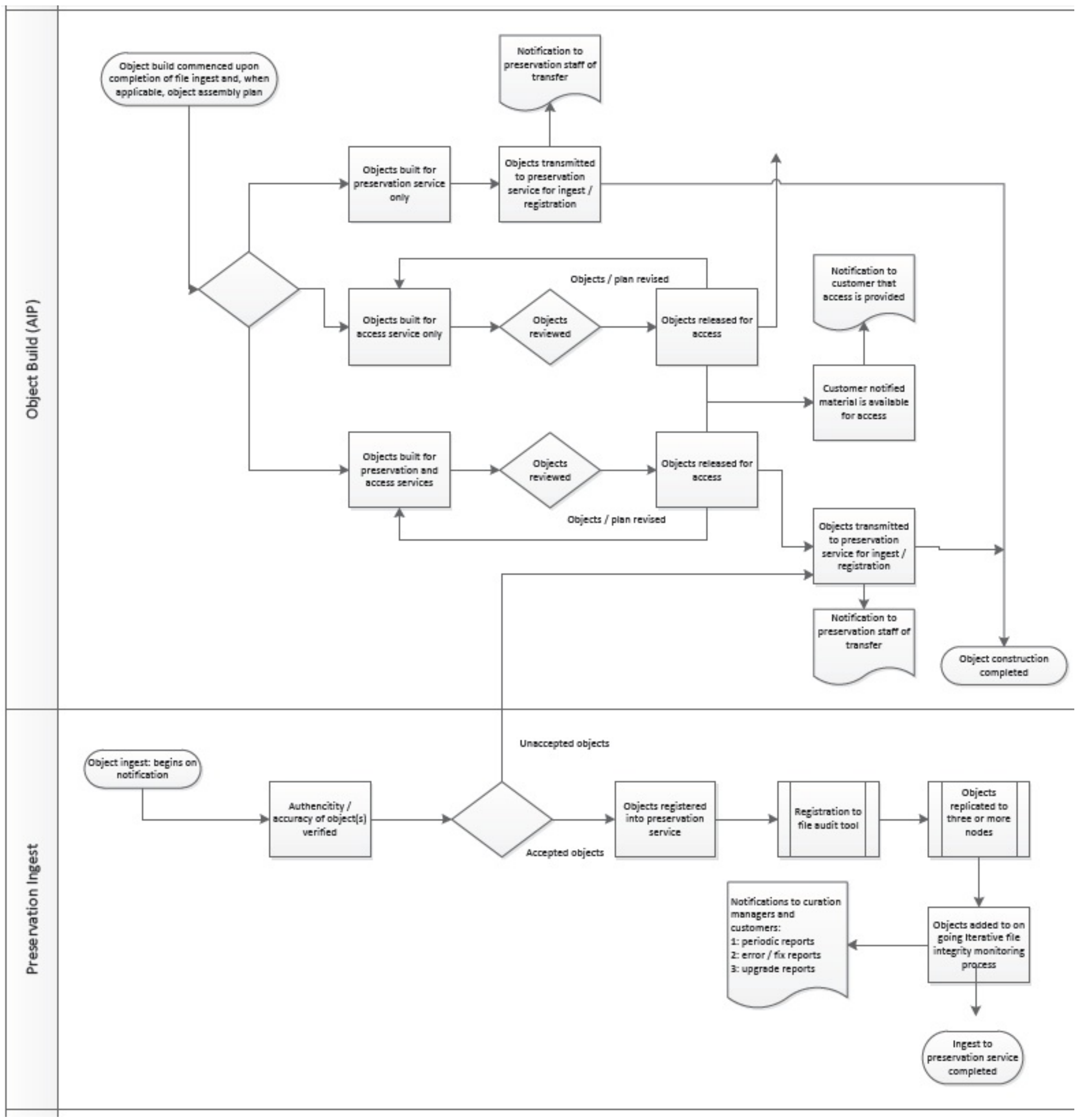
- Utilization of research paper
  - Usage history
  - Metadata enrichment
  - Usage Pattern / Citation
  - Collaboration / sharing
- Data Management
  - Access security
  - Right and privileges
  - Digital Asset Management
  - Stewardship / Data Curation
- Storage Management
  - Tiered storage
  - Archiving
  - Digital preservation
  - Data Storage / centralize vs decentralize
  - Discarding / disposition

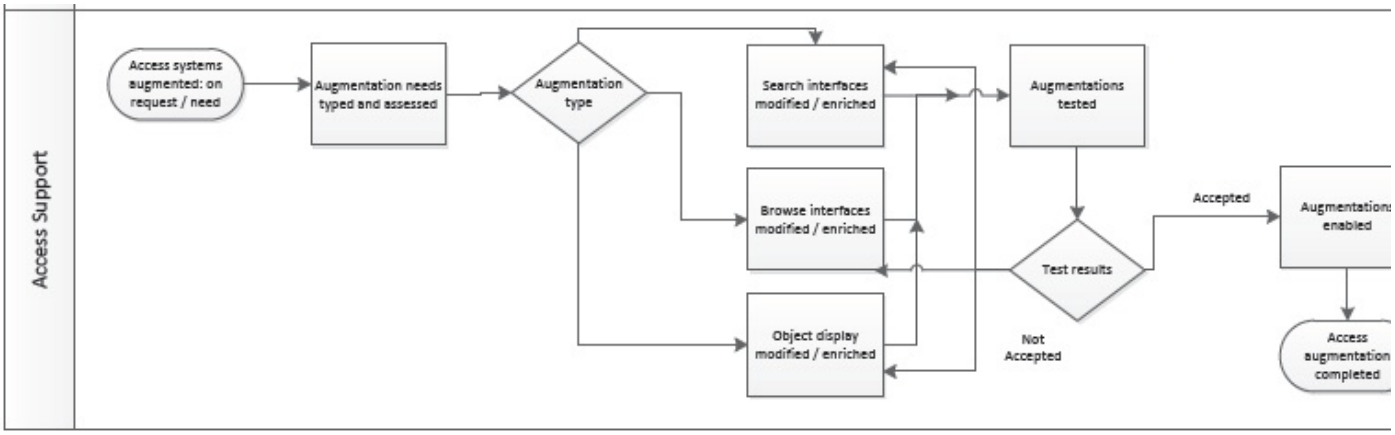
Presented by Imtiaz Khan, IBM, at the National Science Foundation Research Data Lifecycle Management Workshop Princeton, NJ July 18-20, 2011.

26. University of California San Diego Digital Curation Program




Ingest (SIP part)
File Ingest: begins on notification
Ingest of Customer's content files
PID's assigned
Files getting preservation support
Technical Metadata Extracted
Checksum(s) generated
Stub metadata record created with file references
Files getting access support only
Content files and stub records registered in Access Service
Stub metadata record created with file references
Files getting preservation support only
Content files and stub records registered in Preservation Service
Files getting access and preservation support
List of stub records and file identifiers
List of stub records and file identifiers sent to metadata staff, notifying that file ingest is completed
File ingest completed
Metadata (SIP part)
Metadata work: begins on notification after data acquition
Customer's metadata analyzed
Identification of primary metadata to be normalized
Identification of rights metadata to be added
Identification of structural metadata to be added
Establish relationship of metadata to files and stub records
Identification of additional customer metadata to be accommodated
Composition of metadata plan
Review of metadata plan with customer
Acceptance
Object assembly plan created
Object Assembly Plan
Revision
Metadata plan
Assembly plan delivered to build staff
Notification to build staff
Metadata preparation completed


Object Build (AIP)
Object build commenced upon completion of file ingest and, when applicable, object assembly plan
Notification to preservation staff of transfer
Objects built for preservation service only
Objects transmitted to preservation service for ingest / registration
Objects / plan revised
Notification to customer that access is provided
Objects built for access service only
Objects reviewed
Objects released for access
Customer notified material is available for access
Objects built for preservation and access services
Objects reviewed
Objects released for access
Objects / plan revised
Objects transmitted to preservation service for ingest / registration
Notification to preservation staff of transfer
Object construction completed
Preservation Ingest
Unaccepted objects
Object ingest: begins on notification
Authencitity / accuracy of object(s) verified
Objects registered into preservation service
Registration to file audit tool
Objects replicated to three or more nodes
Accepted objects
Notifications to curation managers and customers:
1: periodic reports
2: error / fix reports
3: upgrade reports
Objects added to on going Iterative file integrity monitoring process
Ingest to preservation service completed



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

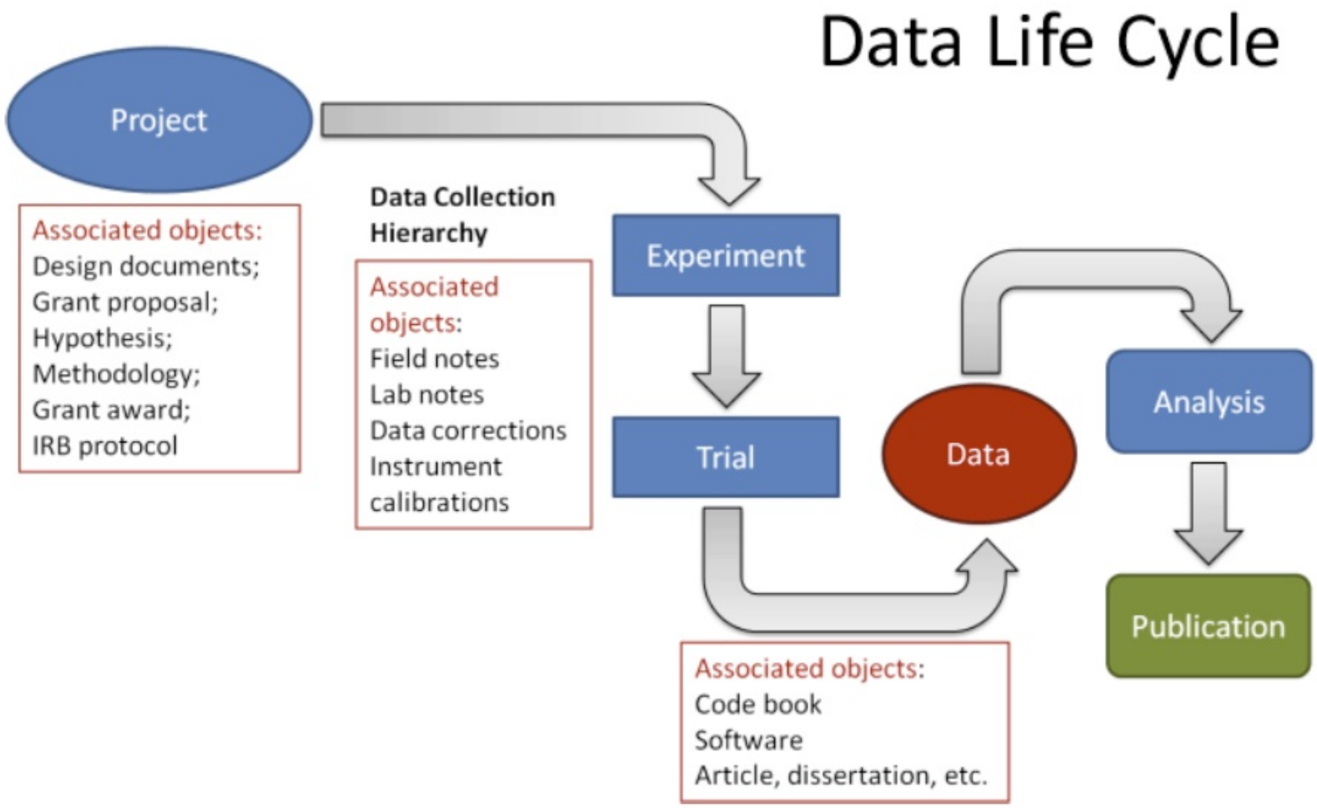
2. "The State of Data"

Data in motion means that data exists in a number of states from its generation to its eventual archiving. Normally, the scientific data life cycle can be described in the following process:

- Data acquisition
- Management and storage
- Processing/modeling
- Post-processing analysis and data mining
- Integration
- Decision Support and Knowledge generation and preservation
- Archiving

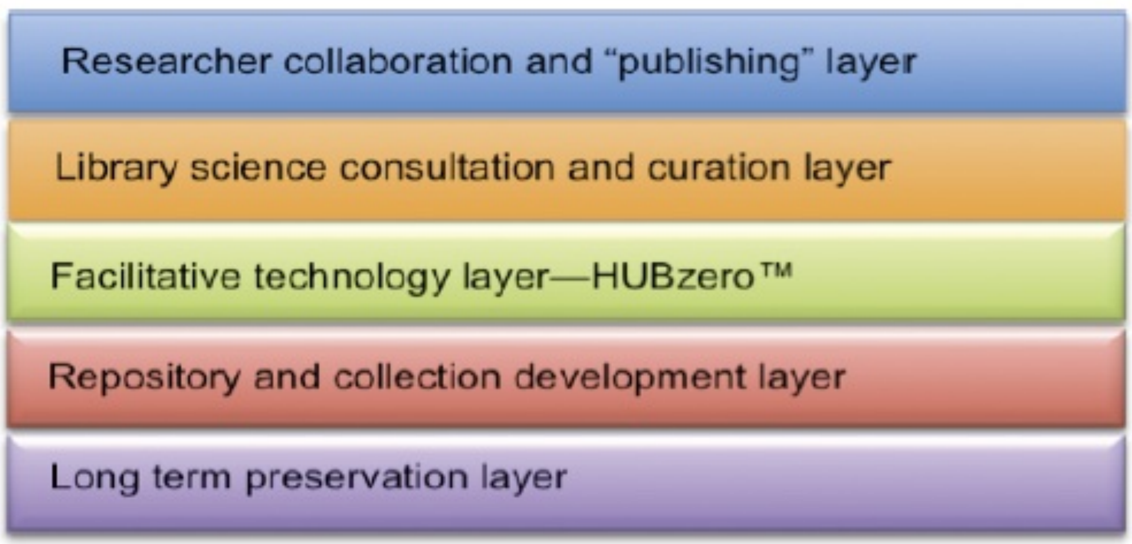
Submitted by Nicholas F. Tsinoremas, Joel Zysman, Christopher Mader, and Jay Blaire, Center for Computational Science, Universtity of Miami, to the National Science Foundation Research Data Lifecycle Management Workshop Princeton, NJ July 18-20, 2011.

28. Managing Research Data Lifecycles through Context



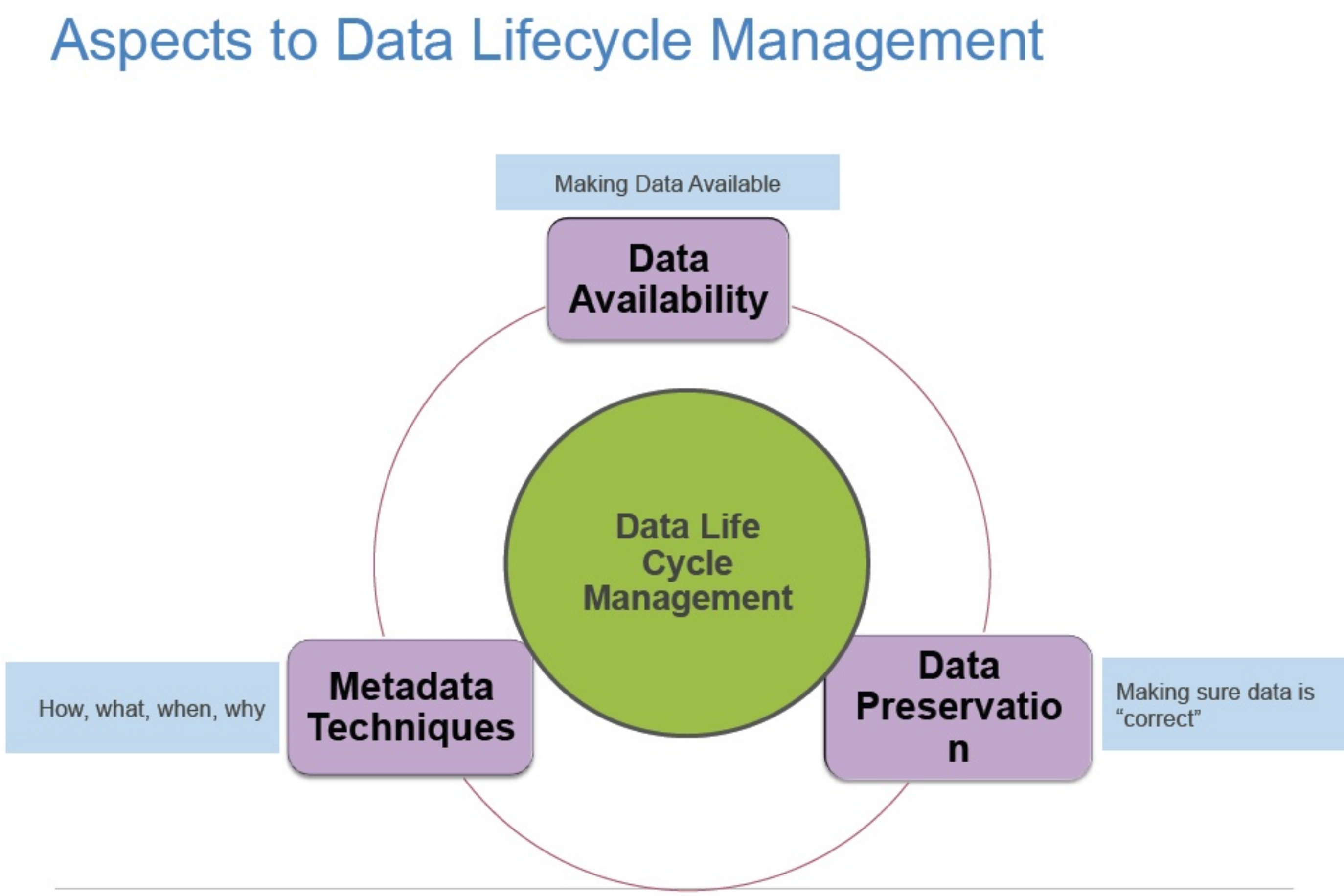
Submitted by Grace Agnew and Ryan Womack, Rutgers University Library, to the National Science Foundation Research Data Lifecycle Management Workshop Princeton, NJ July 18-20, 2011.

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.

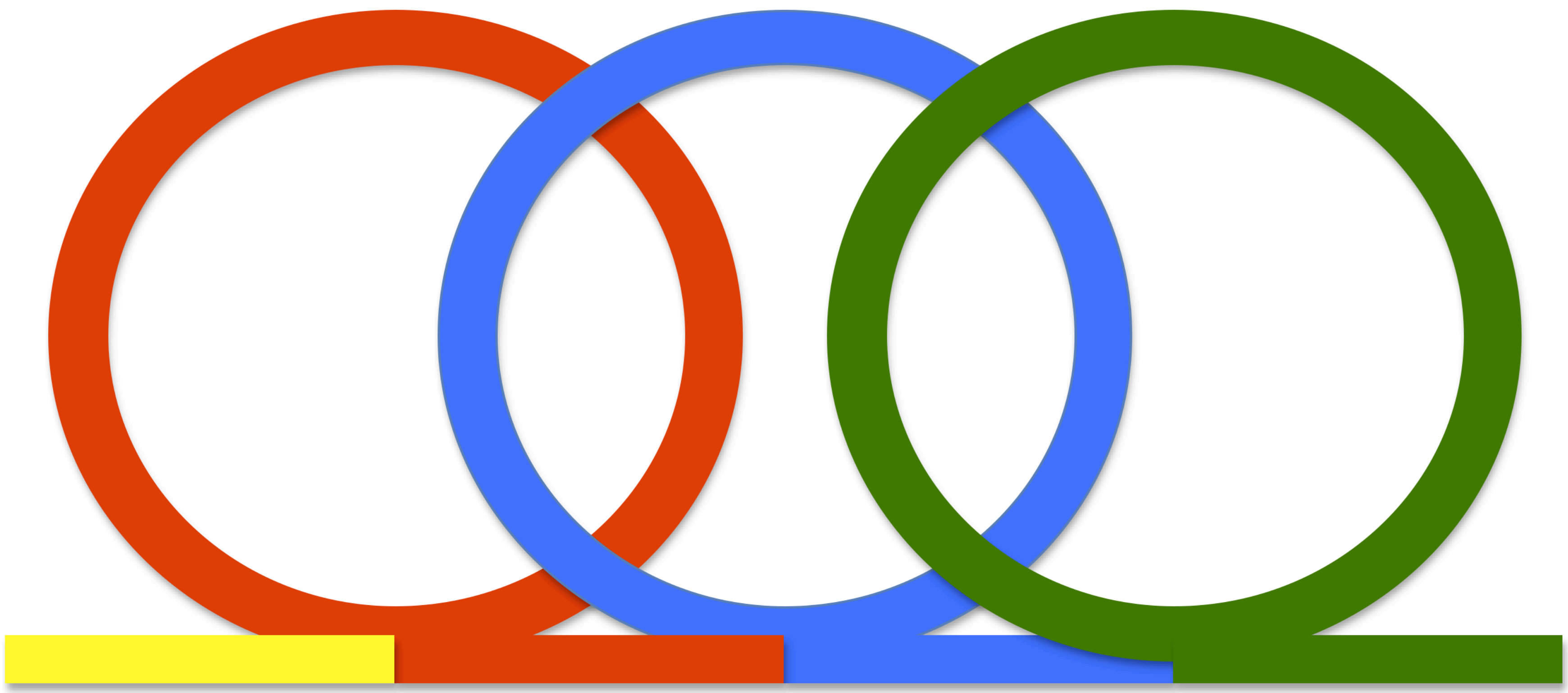
30. Data Lifecycle Management at Dell



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

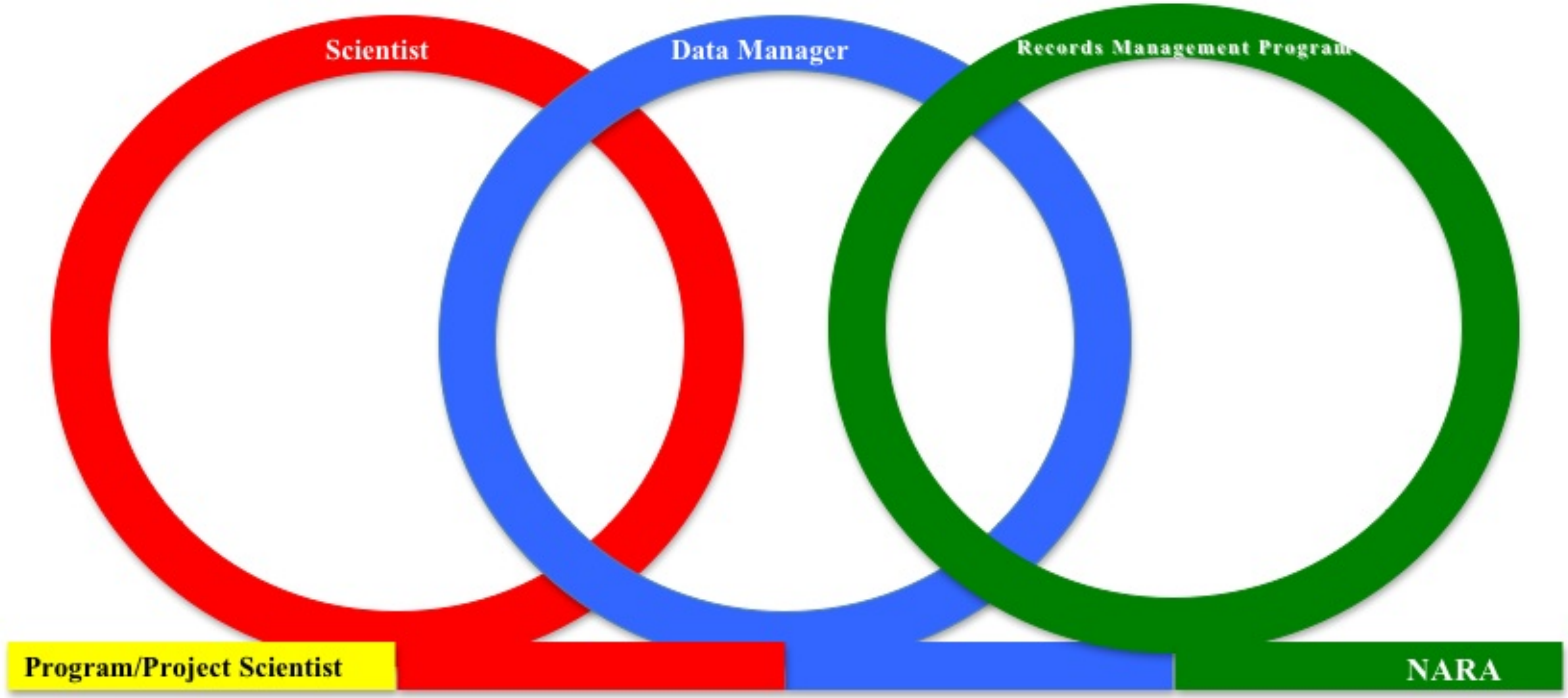
31. John Faundeen & Ellyn Montgomery "Spins"

USGS Data Lifecycle Model
Draft No. 2
7-7-2011

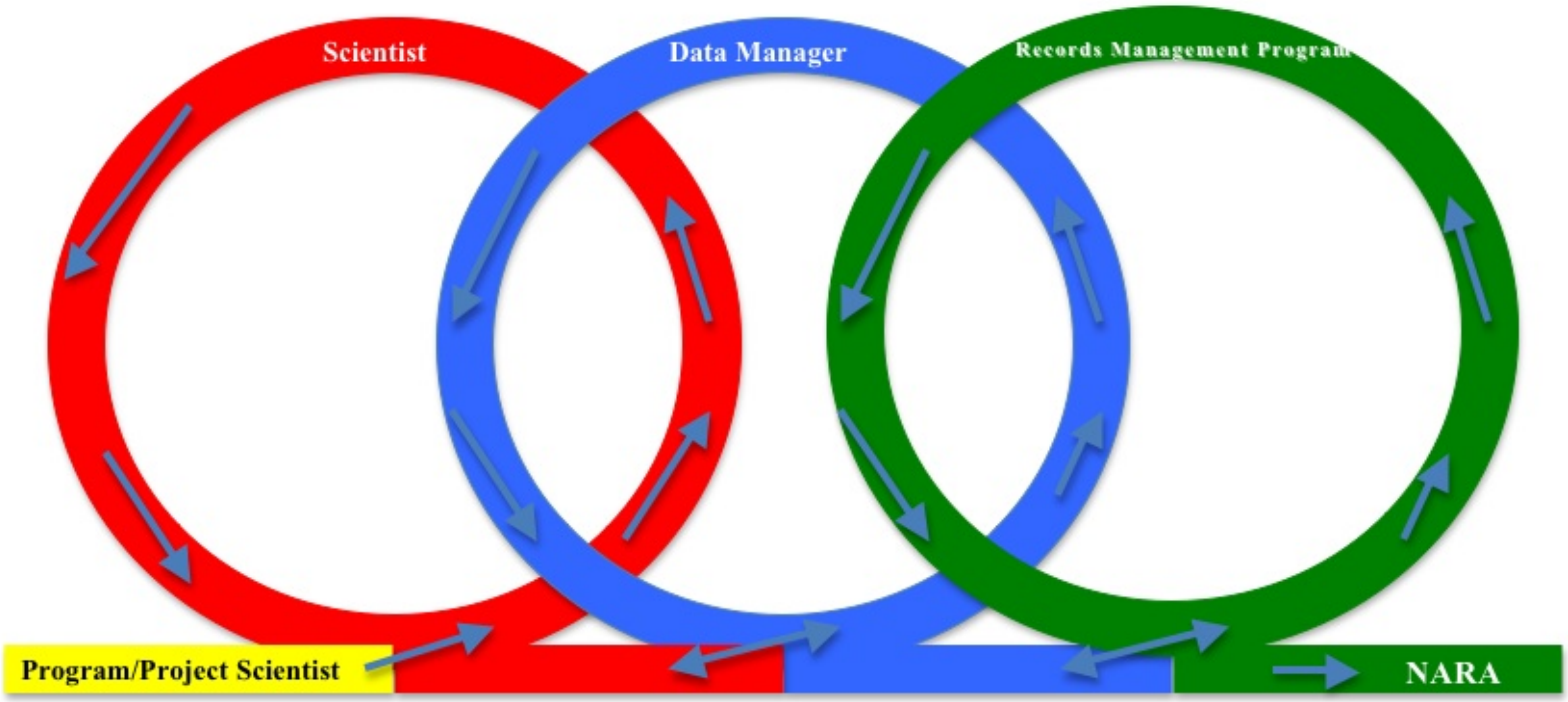


Roles and Timeline

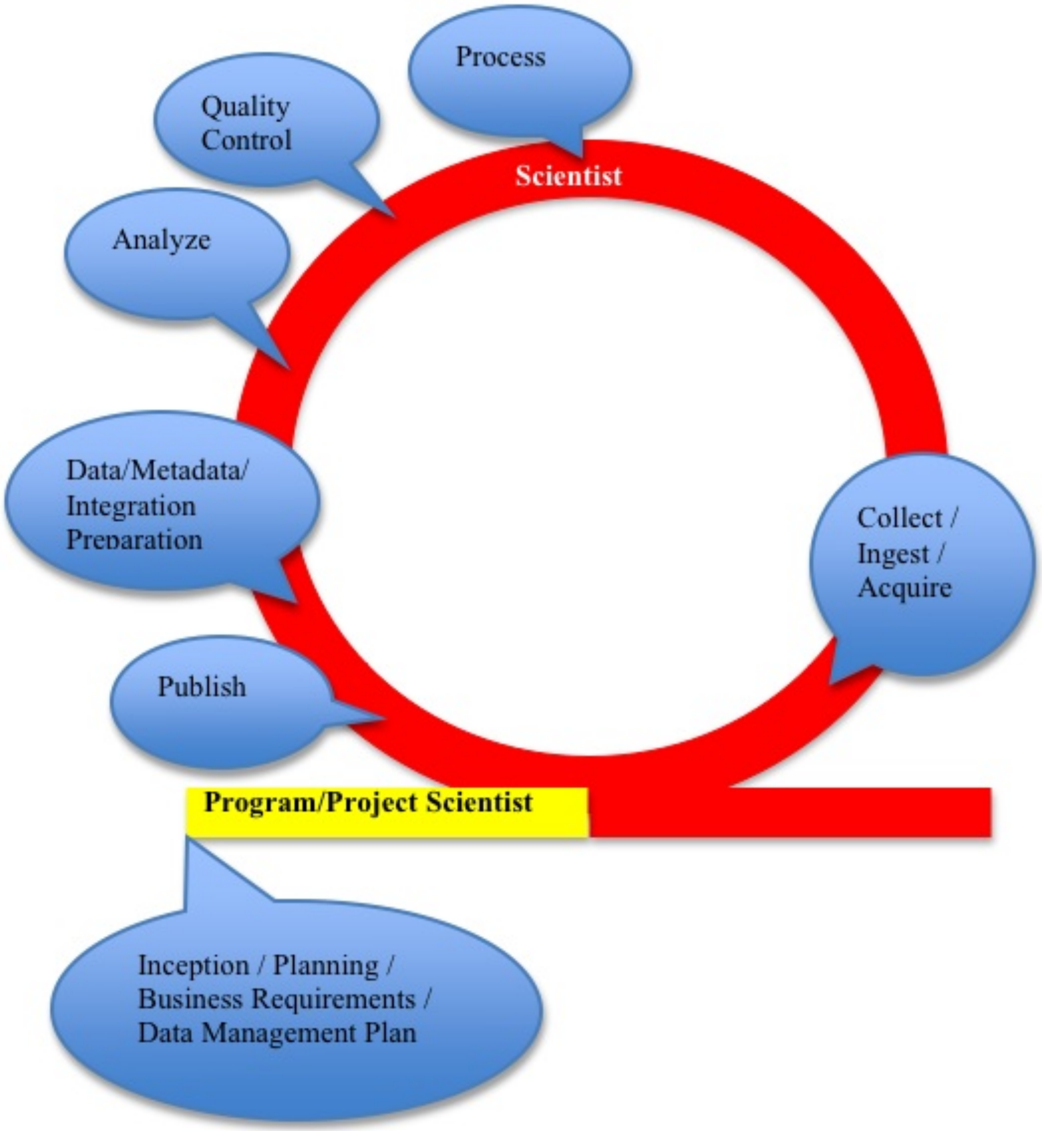
(left to right reading)



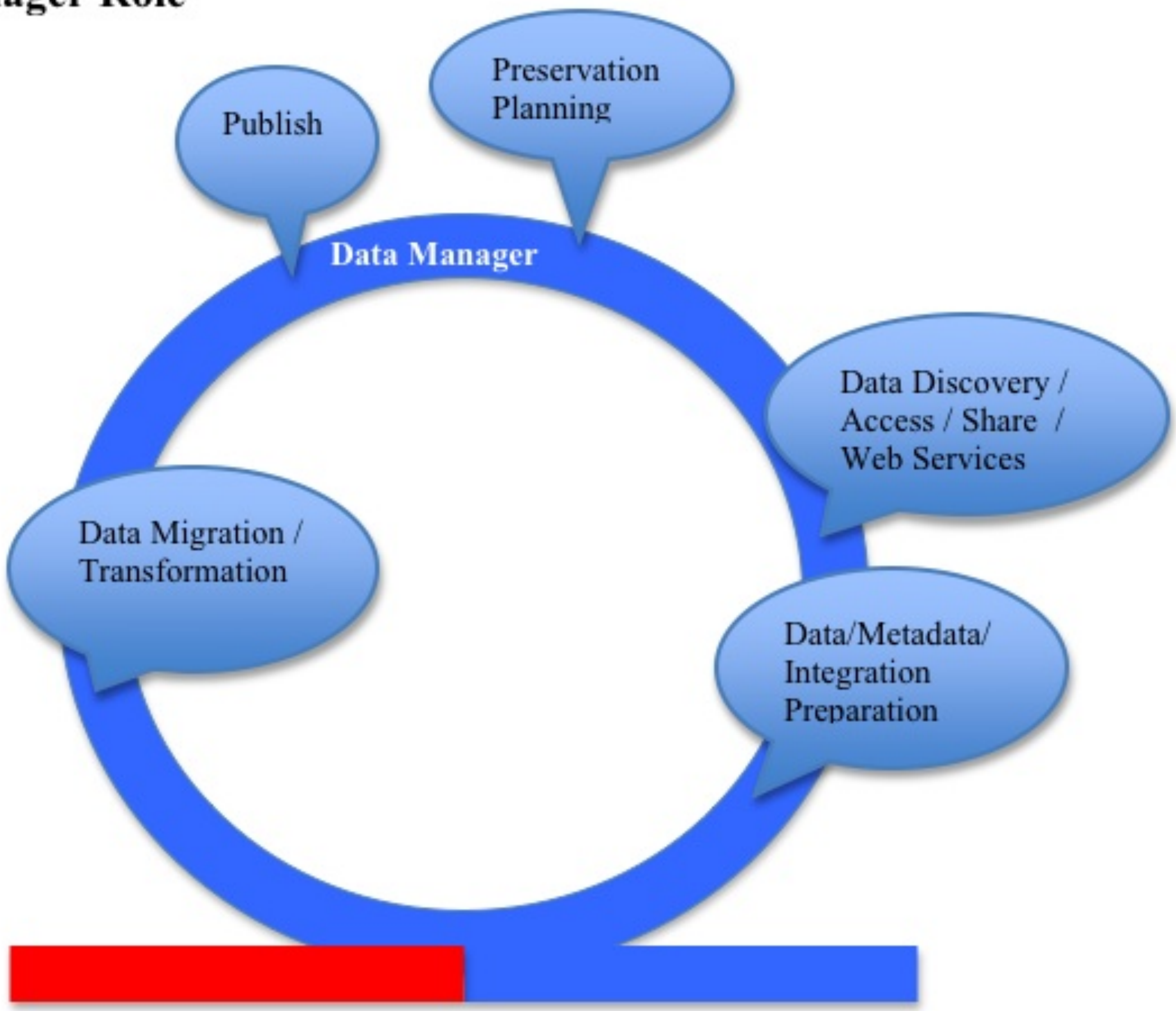
Left to Right Flow With Loop Back Options



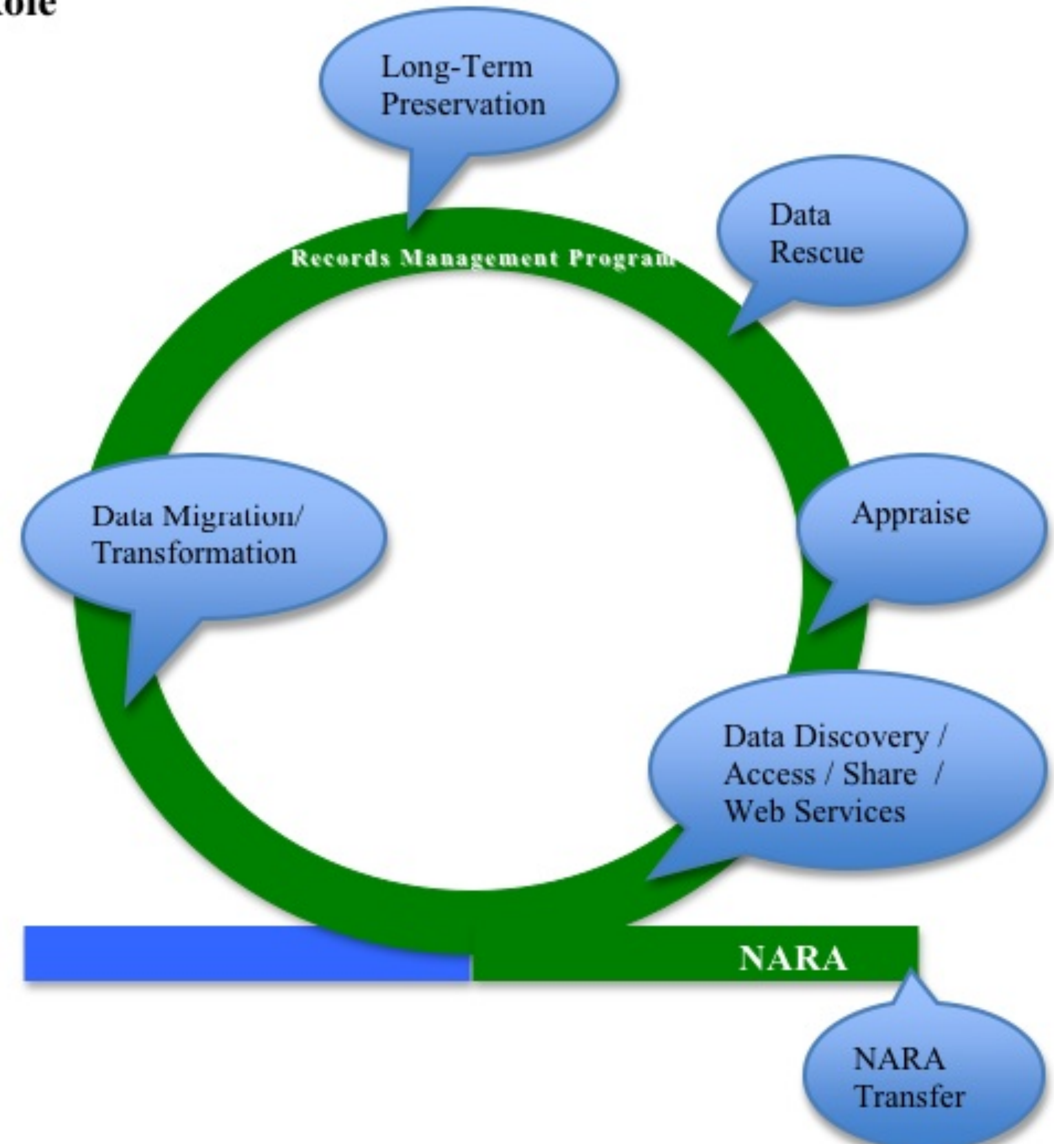
Scientist Role

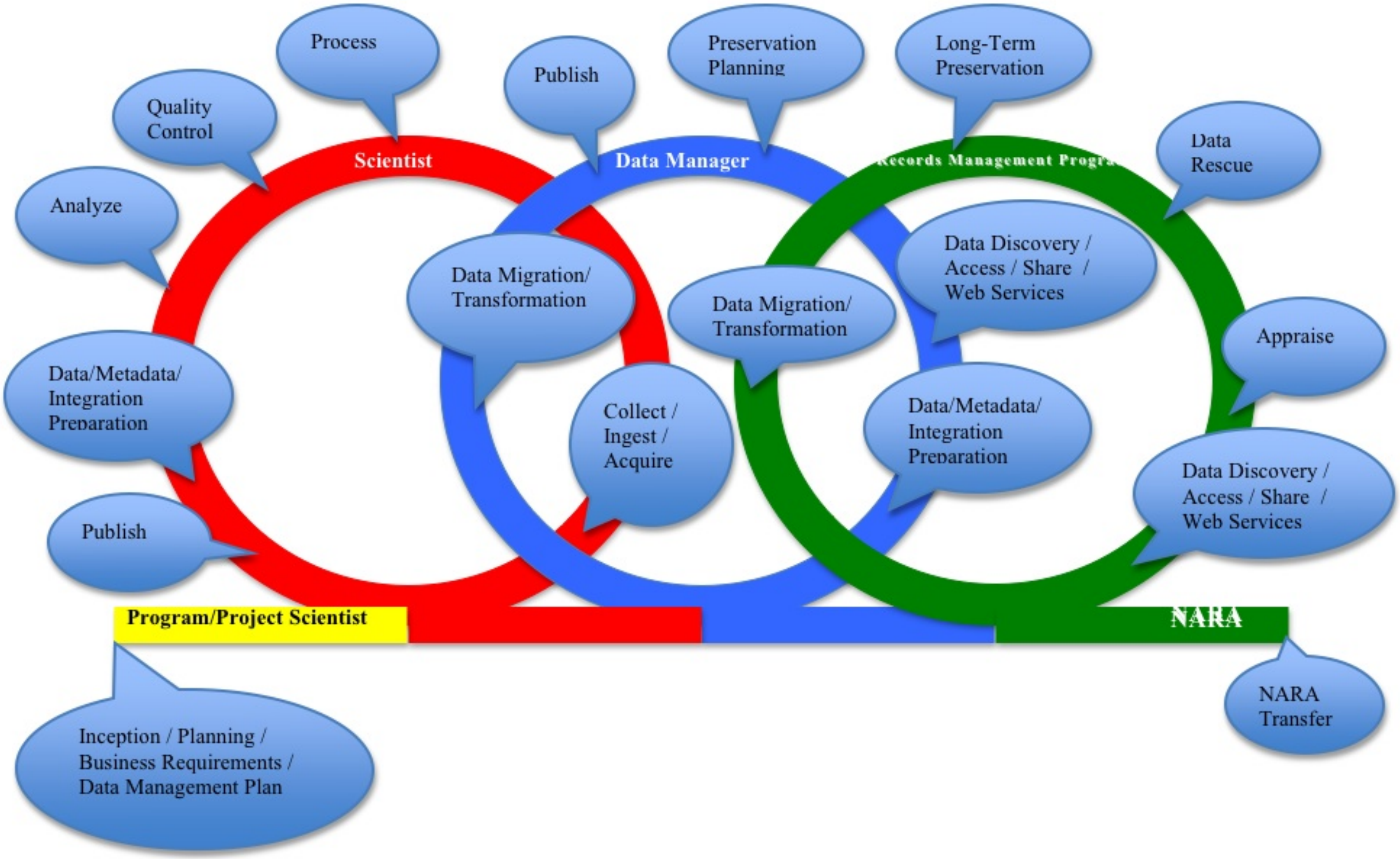


Data Manager Role



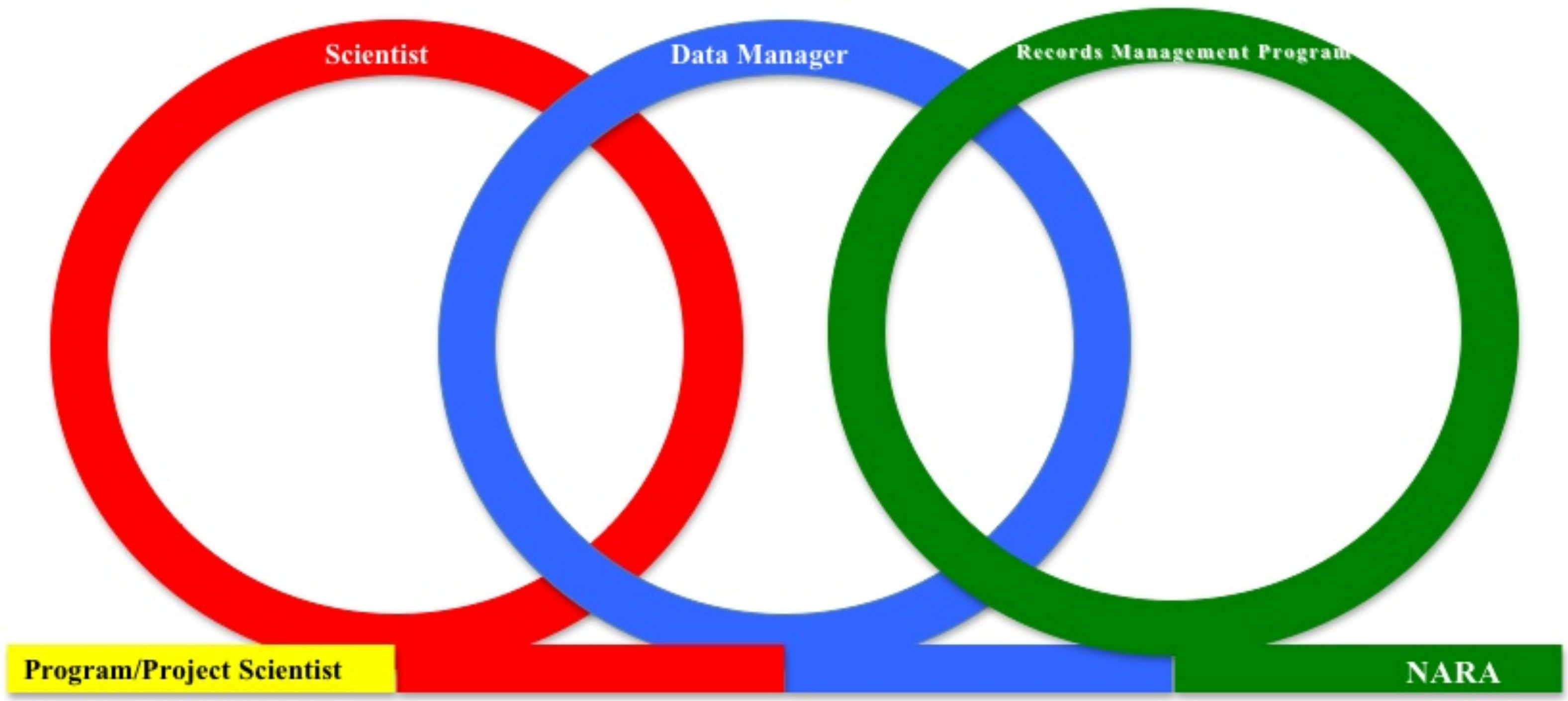
Records Management Program Role





Process
Quality Control
Analyze
Scientist
Publish
Data Manager
Preservation Planning
Long-Term Preservation
Data Rescue
Data Migration/ Transformation
Data Migration/ Transformation
Data Discovery / Access / Share / Web Services
Appraise
Data/Metadata/ Integration Preparation
Collect / Ingest / Acquire
Data/Metadata/ Integration Preparation
Data Discovery / Access / Share / Web Services
Publish
Program/Project Scientist
NARA
NARA Transfer
Inception / Planning / Business Requirements / Data Management Plan

Responsibilities
(lists vs. bubbles with overlaps colored)

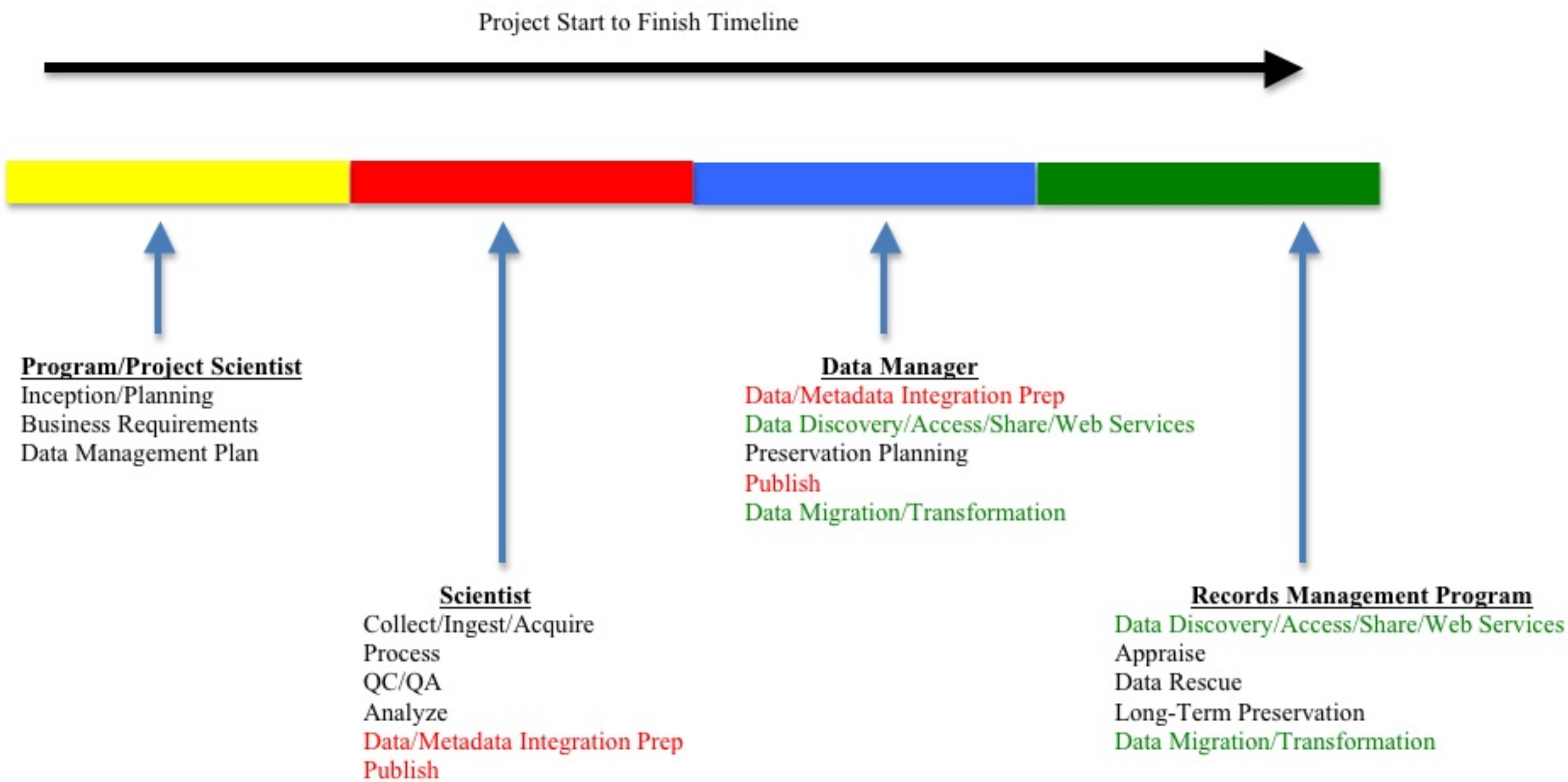


Scientist
Collect/Ingest/Acquire
Process
QC/QA
Analyze
Data/Metadata Integration Prep
Publish

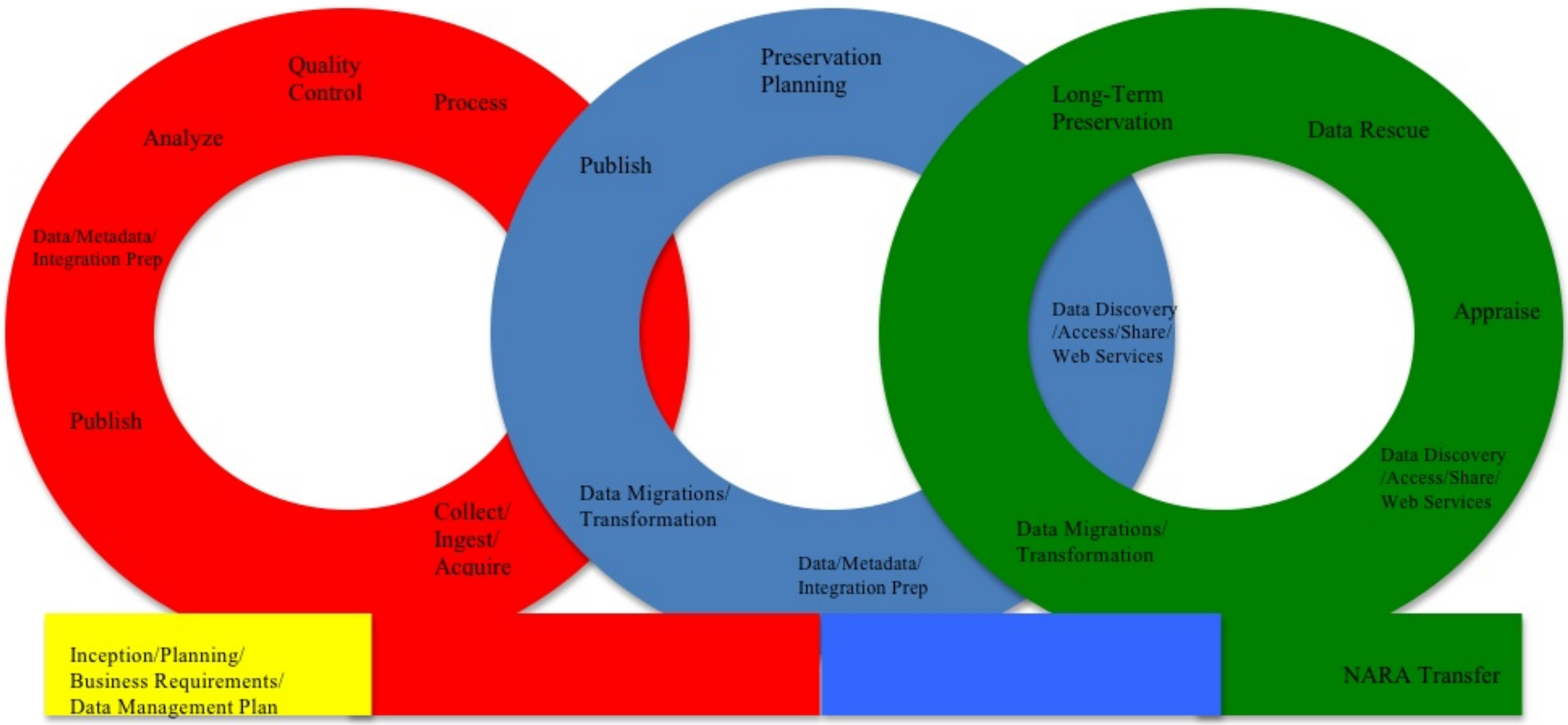
Data Manager
Data/Metadata Integration Prep
Data Discovery/Access/Share/Web Services
Preservation Planning
Publish
Data Migration/Transformation

Records Management Program
Data Discovery/Access/Share/Web Services
Appraise
Data Rescue
Long-Term Preservation
Data Migration/Transformation

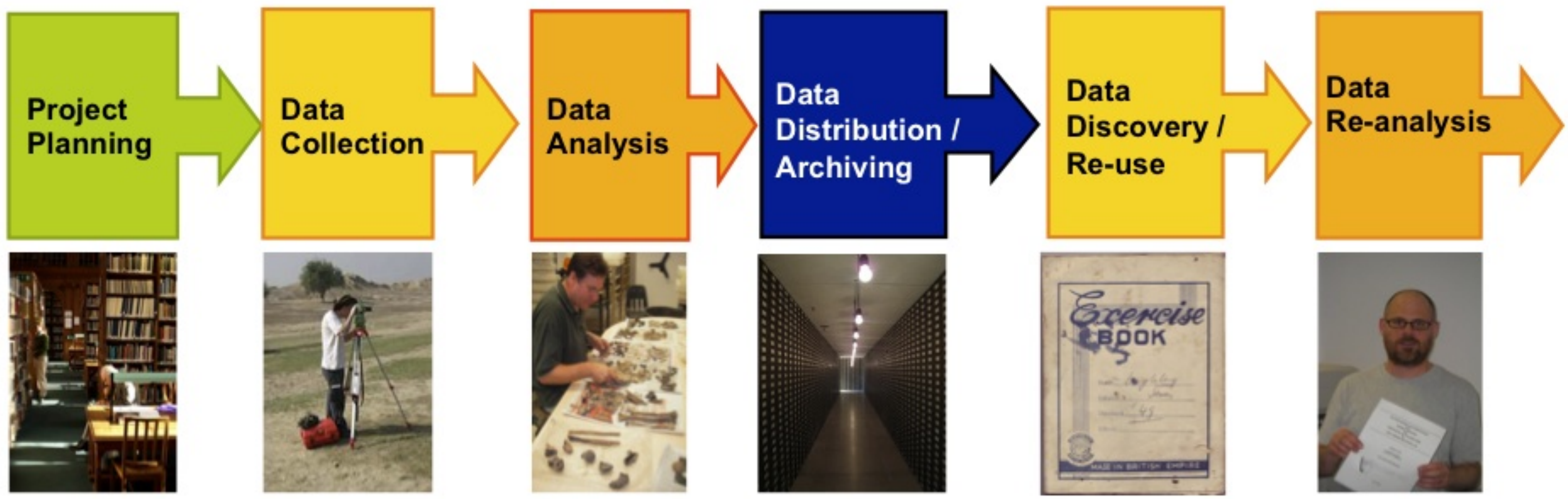
Linear Model – Who Does What When (with overlaps colored)



Ellyn Montgomery "spin"



32. DataTrain Model 1



Open Access Post-Graduate Teaching Materials for
Research Data Management in Archaeology

Created by Lindsay Lloyd-Smith (2011)

Module 2 Data Lifecycles and Management Plans

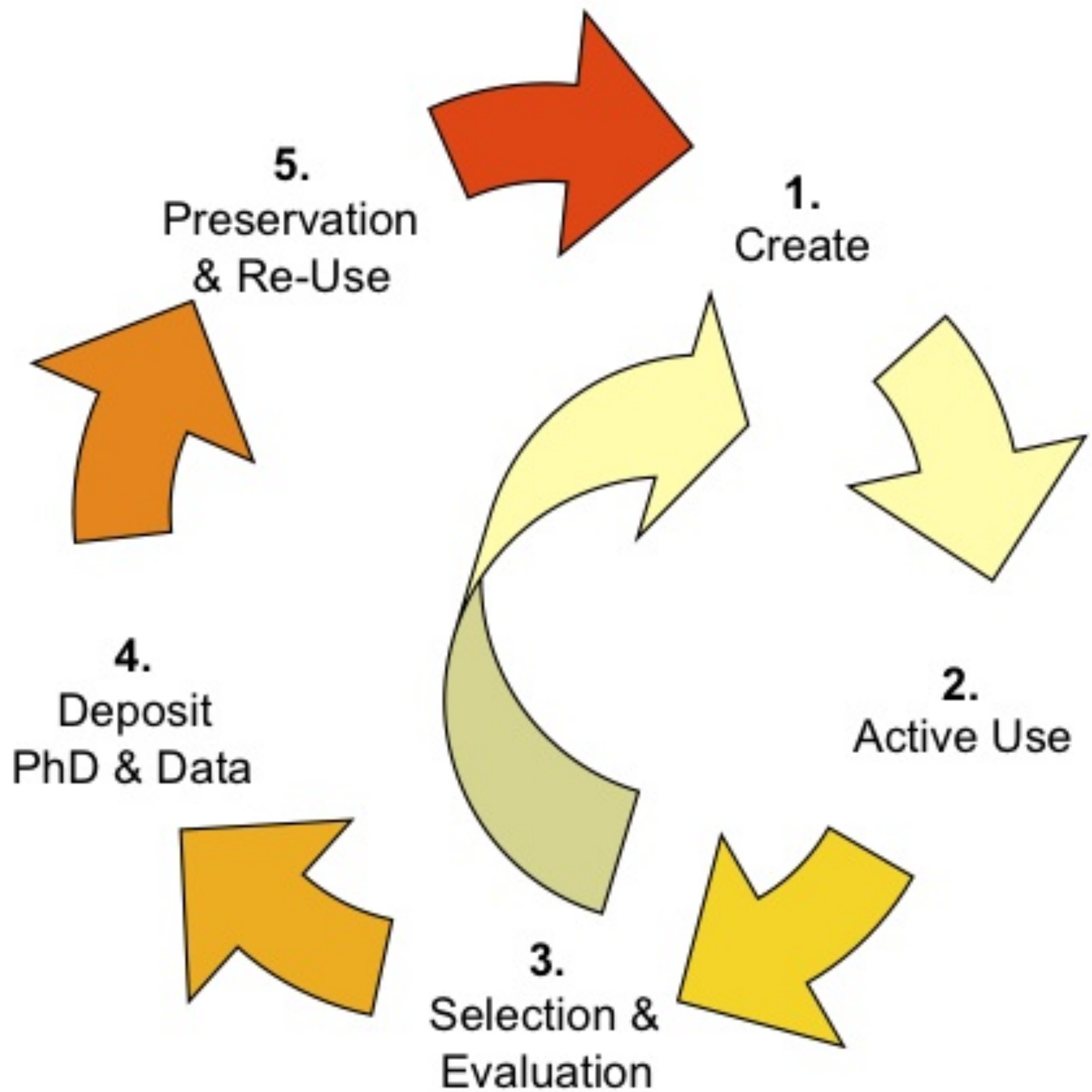
Acknowledgements
This material was created by the JISC-funded DataTrain Project based at the Cambridge University Library.
Project Manager: Elin Stangeland (Cambridge University Library)
Project advisors: Stuart Jeffrey (Archaeology Data Service), Sian Lazar (Department of Anthropology, Cambridge University), Irene Peano (DataTrain Project Officer – Social Anthropology), Cameron Petrie (Department of Archaeology, Cambridge University), Grant Young (Cambridge University Library), and Anna Collins (DSpace@Cambridge Research Data and Digital Curation Officer).

Image credits:
Slide 3 images from left to right:
The Haddon Library courtesy of Cameron Petrie, Department of Archaeology, Cambridge University.
Surveying at Charsadda Pakistan, courtesy of Cameron Petrie, Department of Archaeology, Cambridge University.
John Krigbaum recording human skeletal remains from Niah Cave, Sarawak. Photograph: L. Lloyd-Smith.
Archives des députés allemands: www.flickr.com/photos/hamadryades/2549161782/
Harrisson Excavation Archive field notebook, image courtesy of the Sarawak Museum
Rob Law printing out PhD Thesis. Photograph: L. Lloyd-Smith

Slide 4: DCC Data Lifecycle and image of *Checklist for a Data Management Plan* courtesy of the Digital Curation Centre

UNIVERSITY OF CAMBRIDGE DataTrain ads ARCHAEOLOGY DATA SERVICE JISC

33. DataTrain Model 2



Open Access Post-Graduate Teaching Materials for
Research Data Management in Archaeology

Created by Lindsay Lloyd-Smith (2011)

Module 2 Data Lifecycles and Management Plans

Acknowledgements

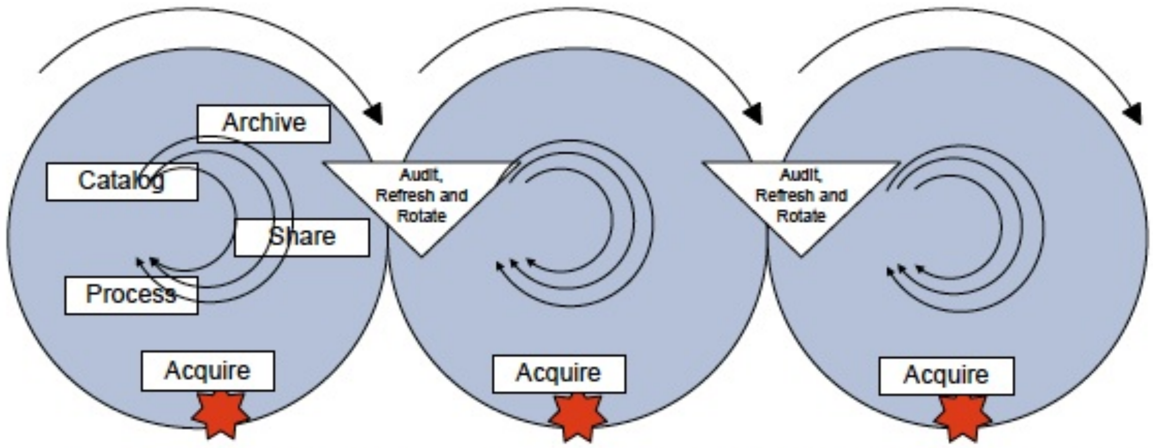
This material was created by the JISC-funded DataTrain Project based at the Cambridge University Library.
Project Manager: Elin Stangeland (Cambridge University Library)
Project advisors: Stuart Jeffrey (Archaeology Data Service), Sian Lazar (Department of Anthropology, Cambridge University), Irene Peano (DataTrain Project Officer – Social Anthropology), Cameron Petrie (Department of Archaeology, Cambridge University), Grant Young (Cambridge University Library), and Anna Collins (DSpace@Cambridge Research Data and Digital Curation Officer).

Image credits:

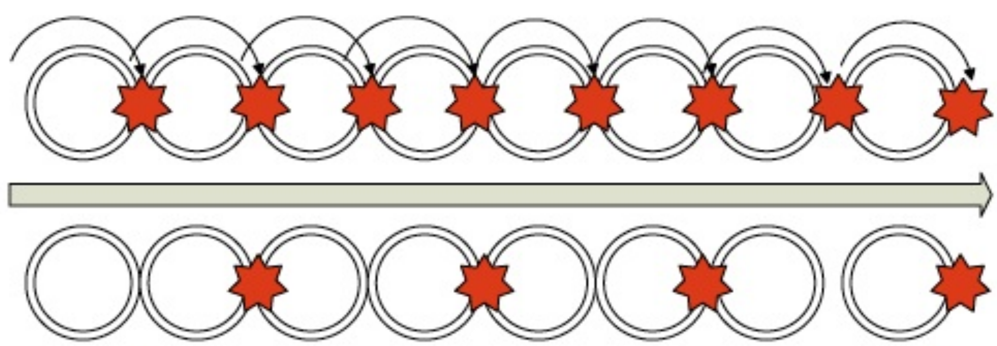
Slide 3 images from left to right:
The Haddon Library courtesy of Cameron Petrie, Department of Archaeology, Cambridge University.
Surveying at Charsadda Pakistan, courtesy of Cameron Petrie, Department of Archaeology, Cambridge University.
John Krigbaum recording human skeletal remains from Niah Cave, Sarawak. Photograph: L. Lloyd-Smith.
Archives des députés allemands: www.flickr.com/photos/hamadryades/2549161782/
Harrisson Excavation Archive field notebook, image courtesy of the Sarawak Museum
Rob Law printing out PhD Thesis. Photograph: L. Lloyd-Smith

Slide 4: DCC Data Lifecycle and image of *Checklist for a Data Management Plan* courtesy of the Digital Curation Centre

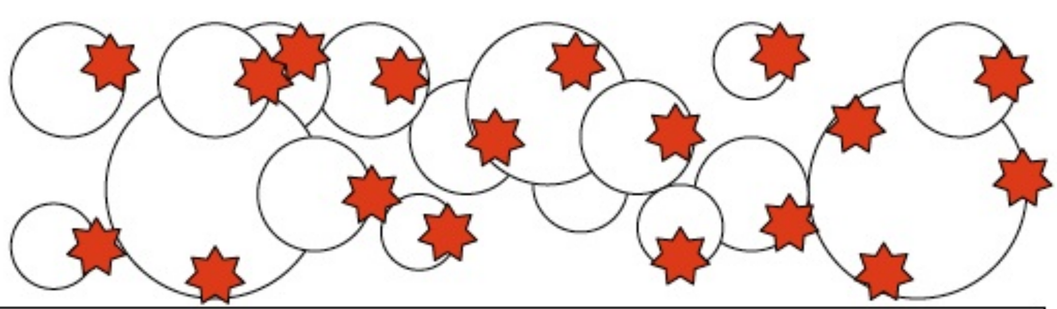
34. Steve Tessler Data Management Cycles



Preservation Data Management Cycles – semiannual audit, refresh, media rotation
Continuous acquisition and archive activities

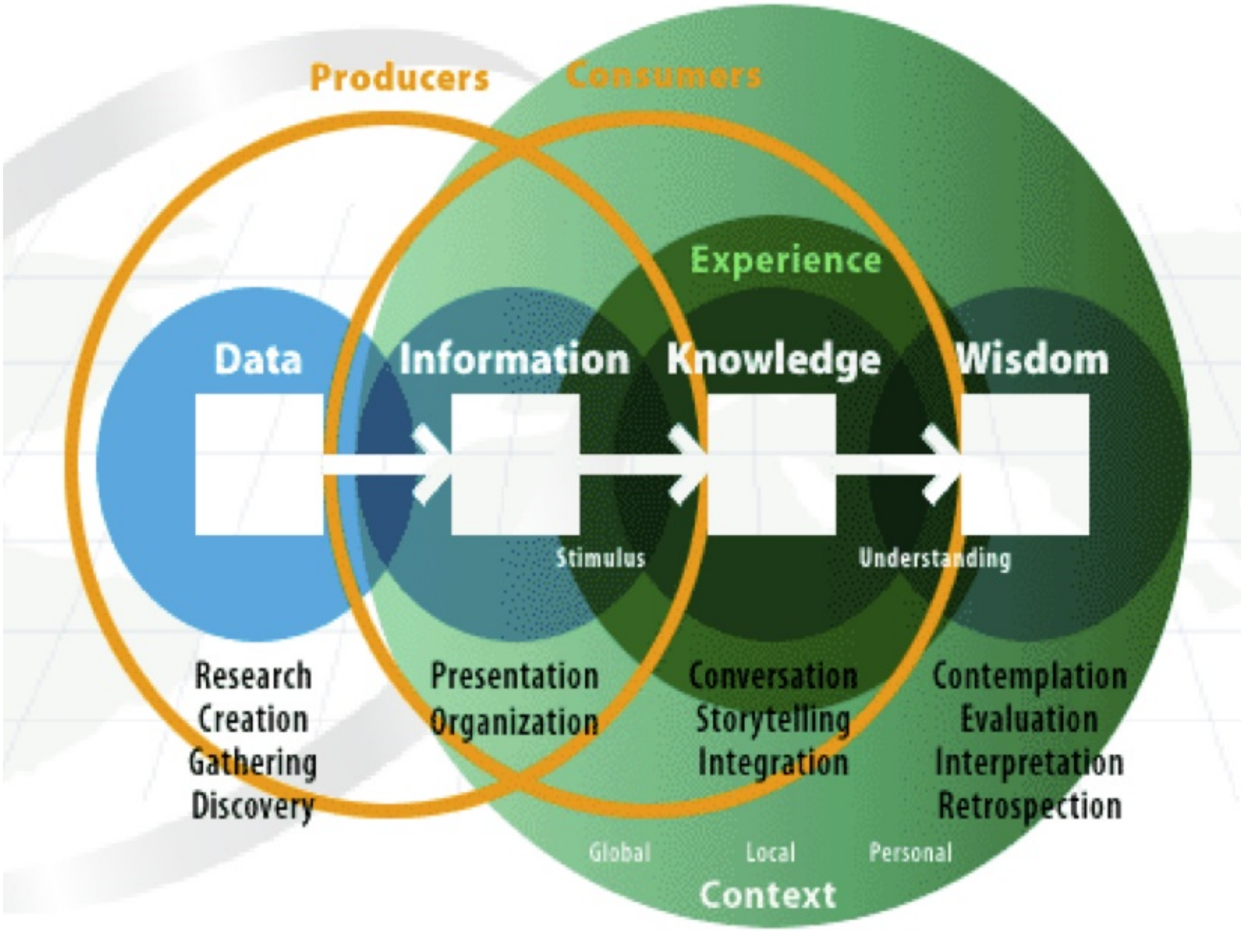


Multi-agency, Long-term Monitoring - Annual Research Data Management Cycles
Hand-off to Preservation cycle is Scheduled for annual/semiannual DM cycle



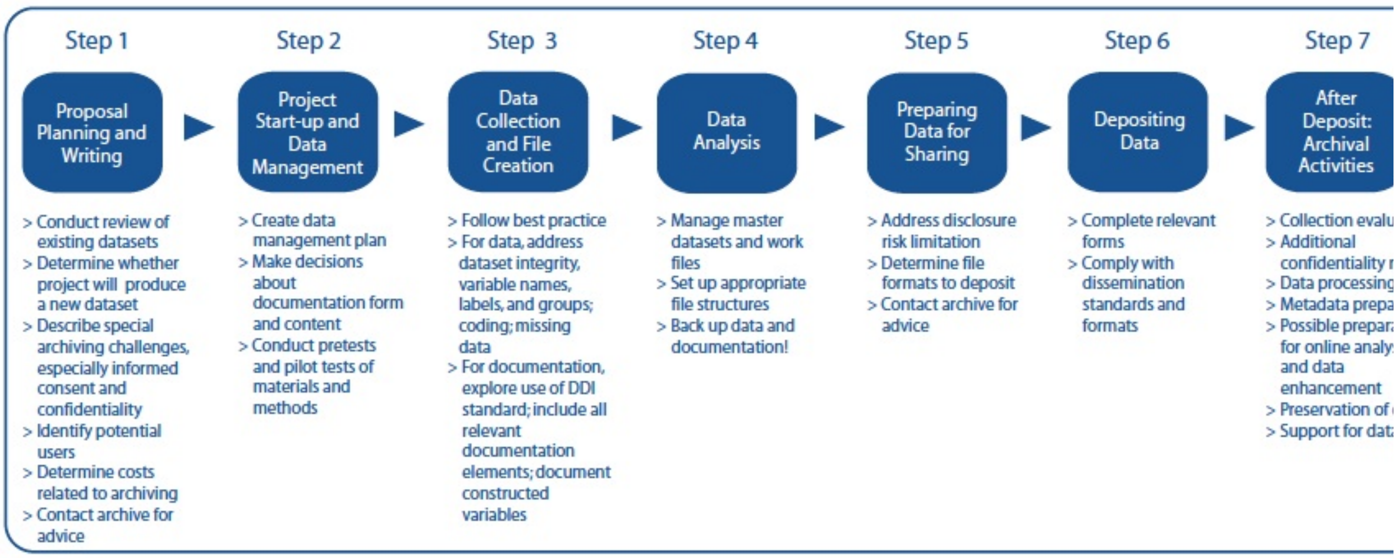
Independent, Short-term Research Project Data Management Cycles (1-5 years)
Hand-off to Preservation cycle at end of project or at multi-year project set-points

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

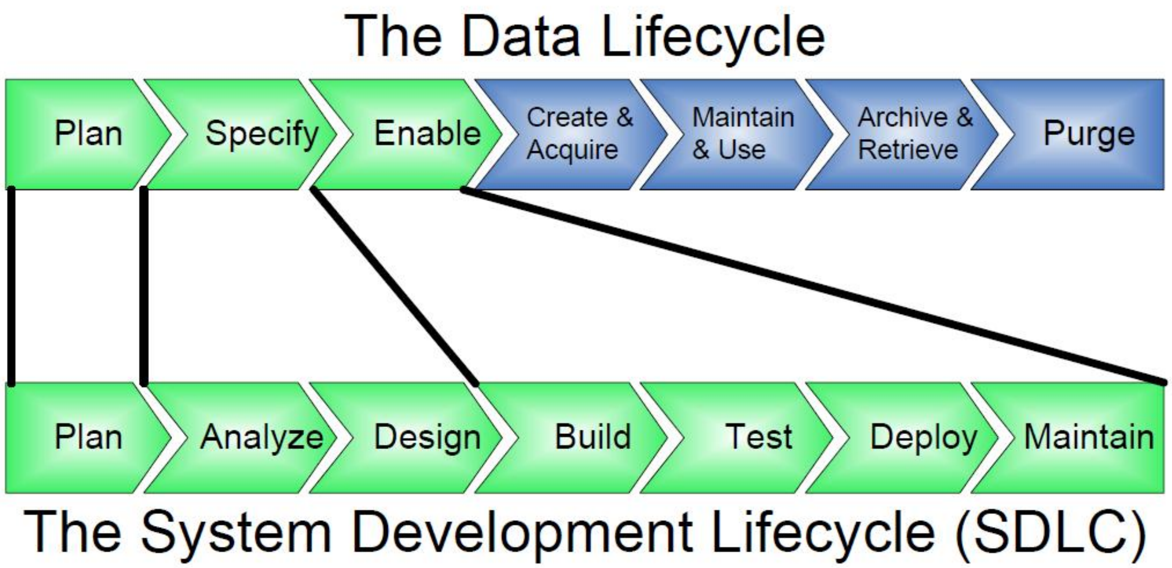
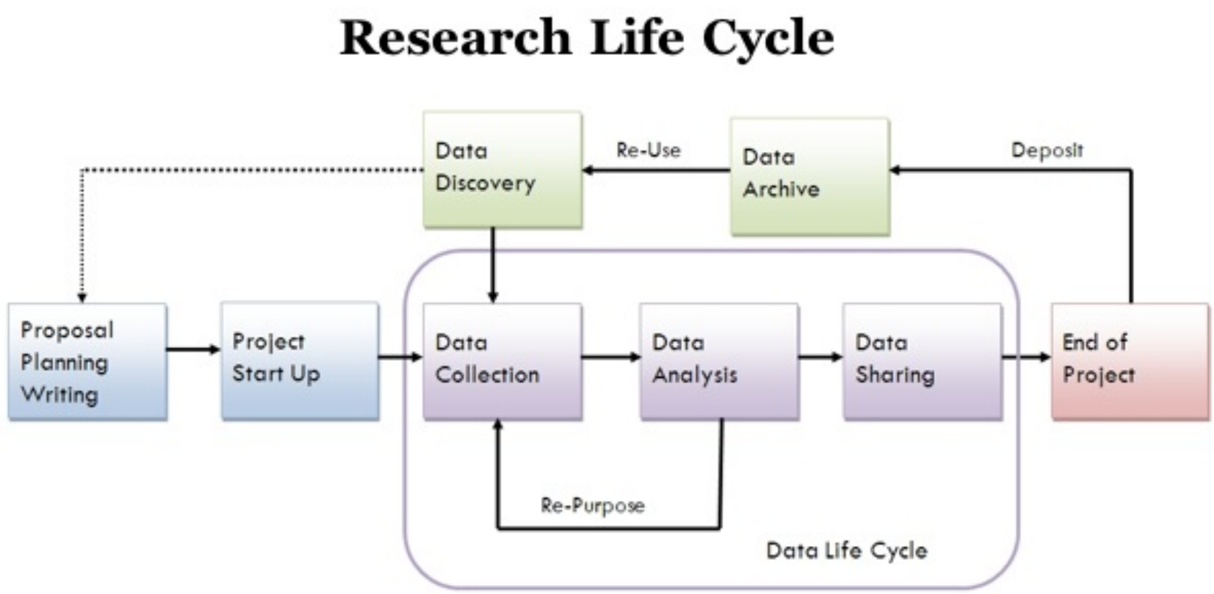


Figure 1.2 The Data Lifecycle and the System Development Lifecycle

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. But really the ideas collected are generally applicable to the kind of scientific projects that uses the CSCS computational and visualization services.

Problems and opportunities

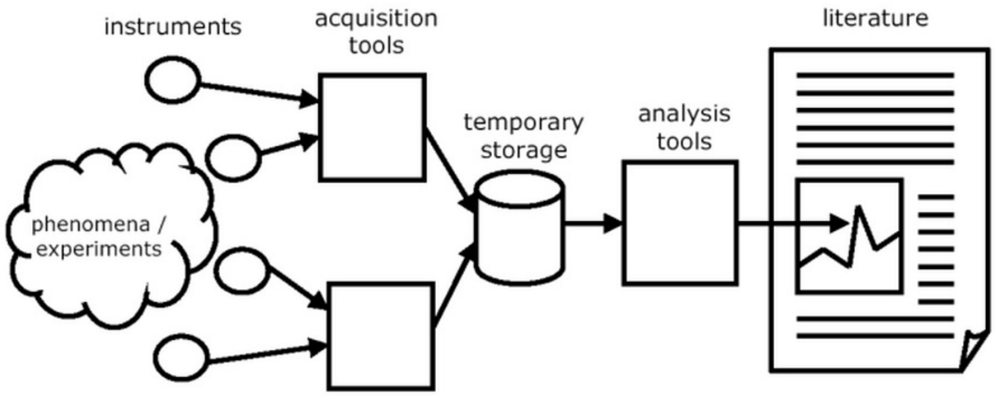
Problems that can be found in current scientific projects are for example:

- 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 and USA. The idea is summarized below:

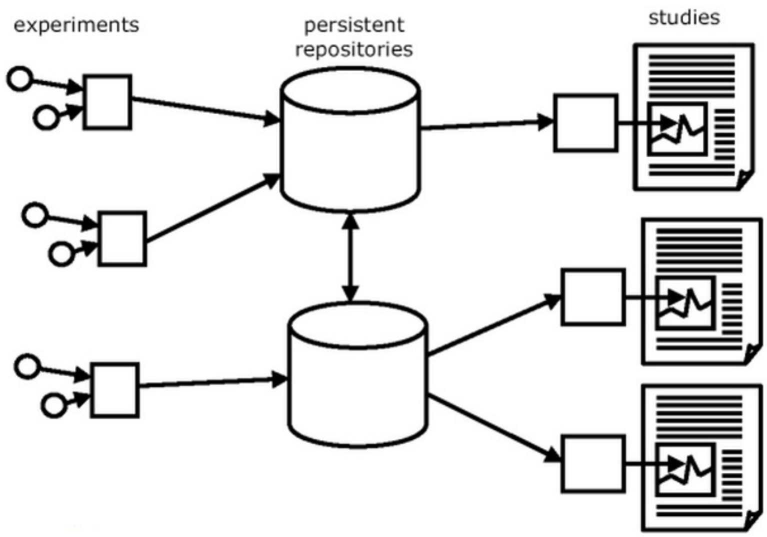
Scientific Data Lifecycle

From this (publish and forget)…



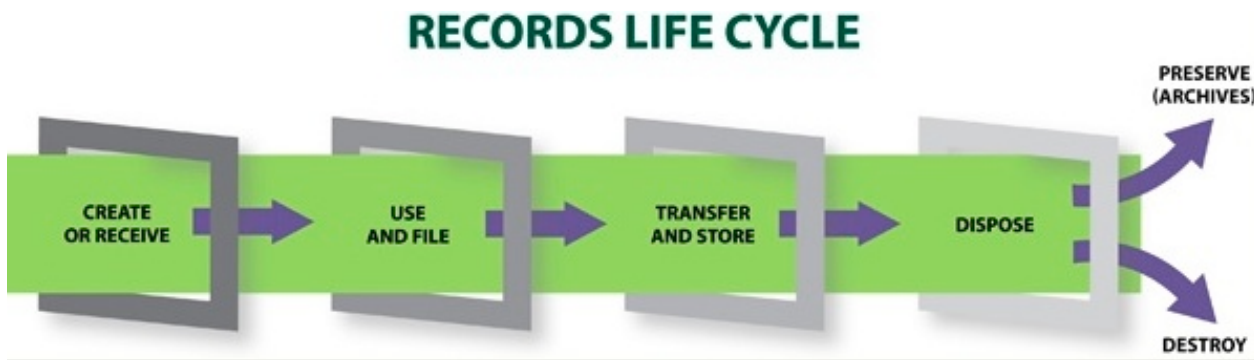
Scientific Data Lifecycle

…to this one



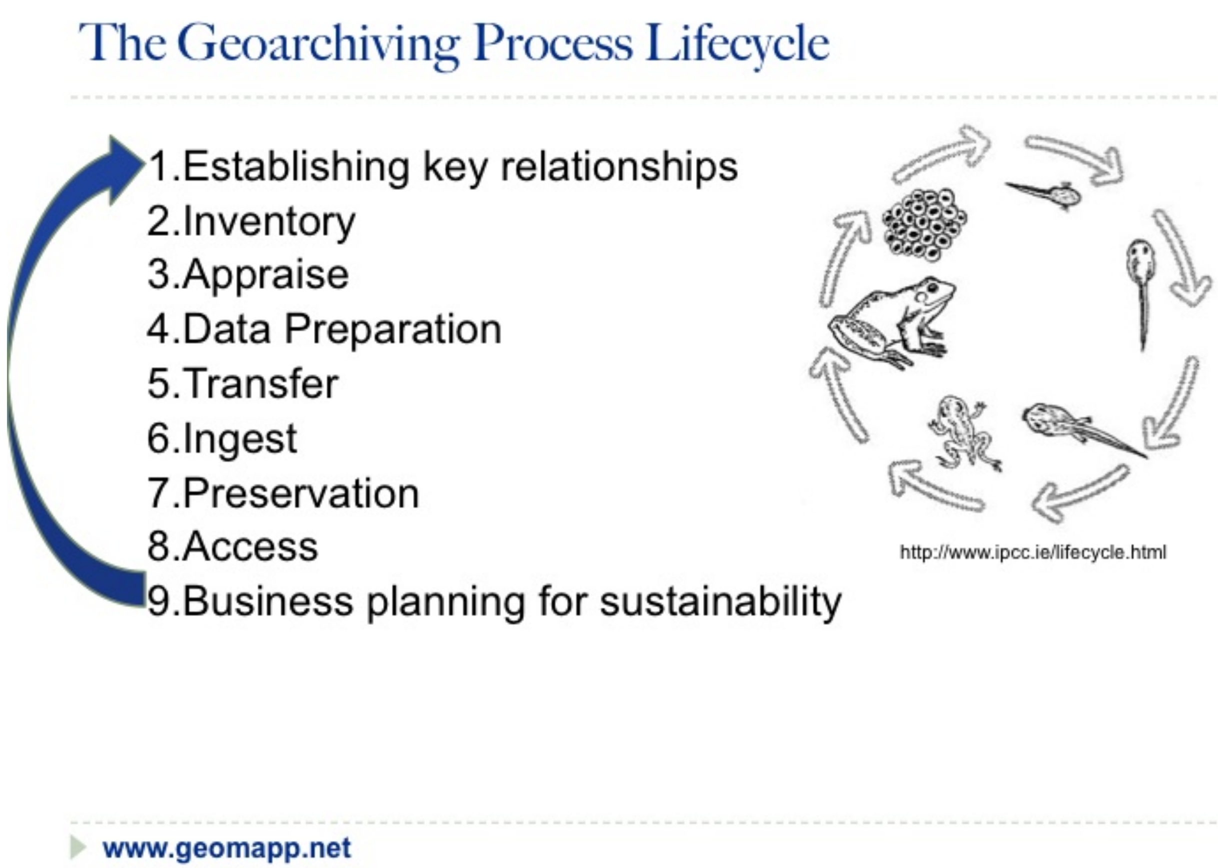
Obtained 8-12-2011 from Mario Valle Web > Scientific Data Management > Introduction URL http://personal.cscs.ch/~mvalle/sdm/scientific-data-management.html

40. Michigan State University Records Life Cycle Model



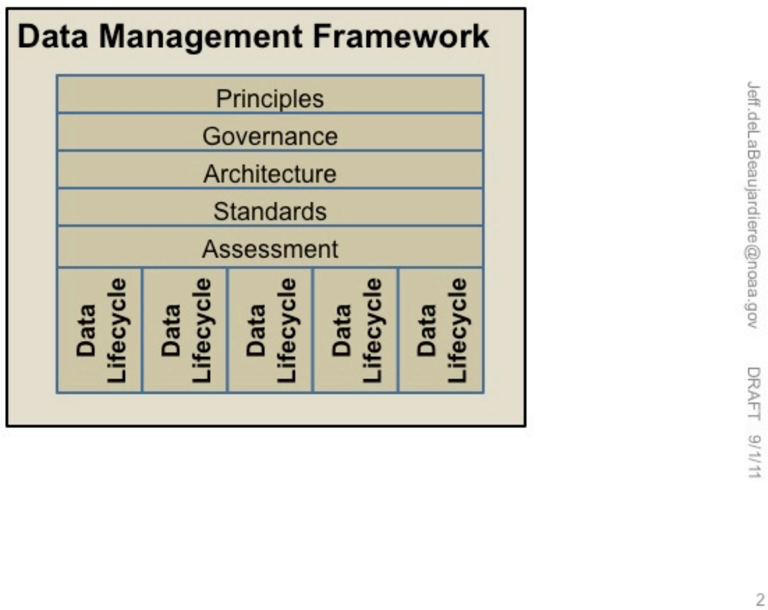
Obtained 8-30-2011 from Michigan State University Archives and Historical Records URL http://archives.msu.edu/records/

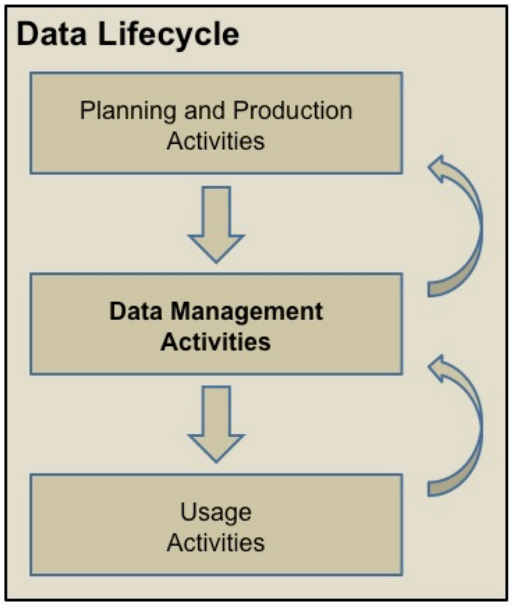
41. GeoMAPP Geoarchiving Process Lifecycle



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.

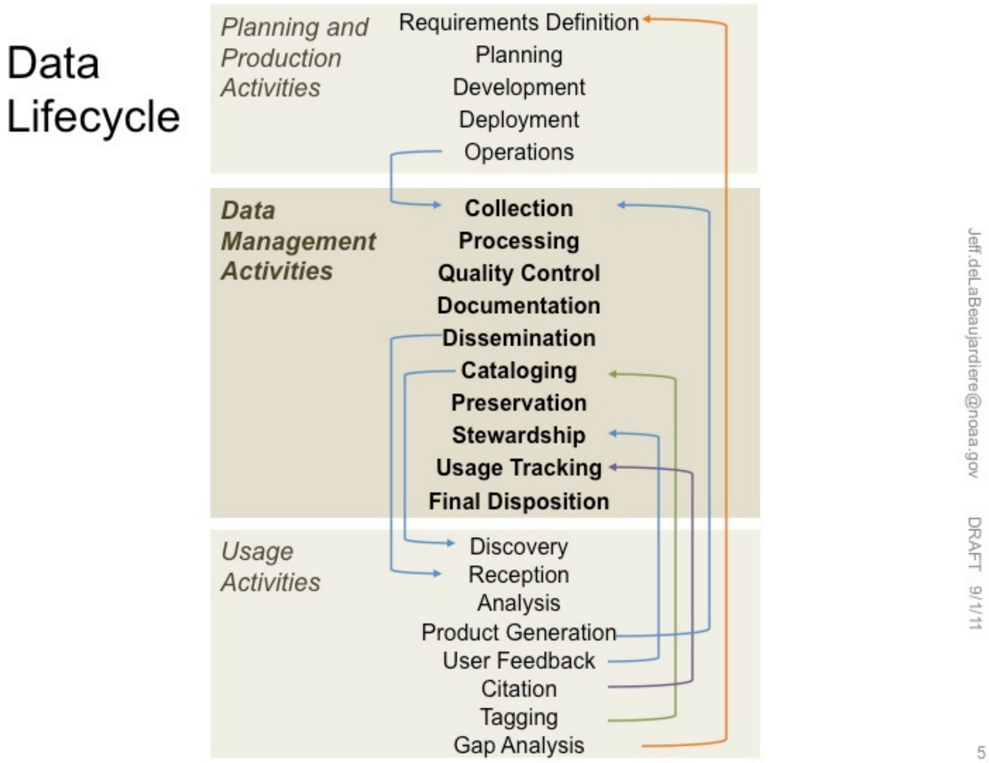
42. Jeff de La Beaujardière's proposed Data Management Framework



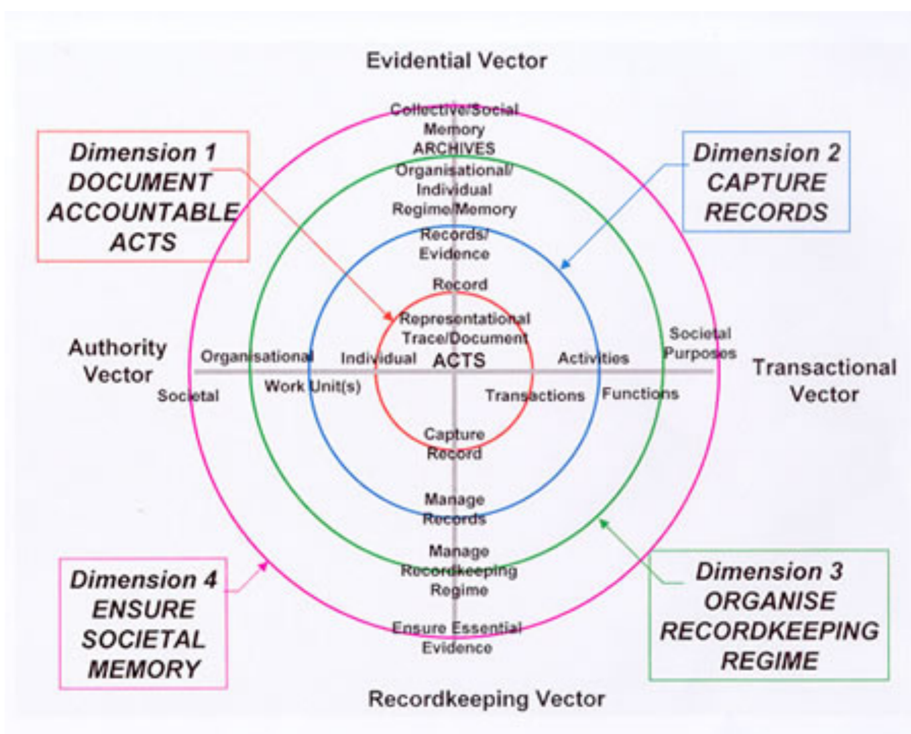


Data Lifecycle

Planning and Production Activities	Requirements Definition Planning Development Deployment Operations
Data Management Activities	**Collection** **Processing** **Quality Control** **Documentation** **Dissemination** **Cataloging** **Preservation** **Stewardship** **Usage Tracking** **Final Disposition**
Usage Activities	Discovery Reception Analysis Product Generation User Feedback Citation Tagging Gap Analysis

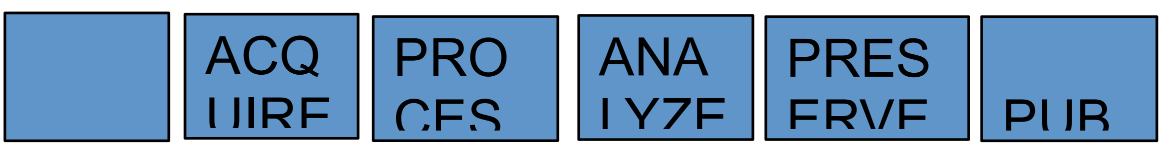


43. Sarah Demb Continuumm Model



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)