

## **Data Lifecycle Models and Concepts**

This is a compilation of data lifecycle models and concepts assembled in part to fulfill Committee on Earth Observation Satellites Working Group on Information Systems and Services and the U.S. Geological Survey Council for Data Integration needs. It is intended to be a living document, which will evolve as new information is discovered.

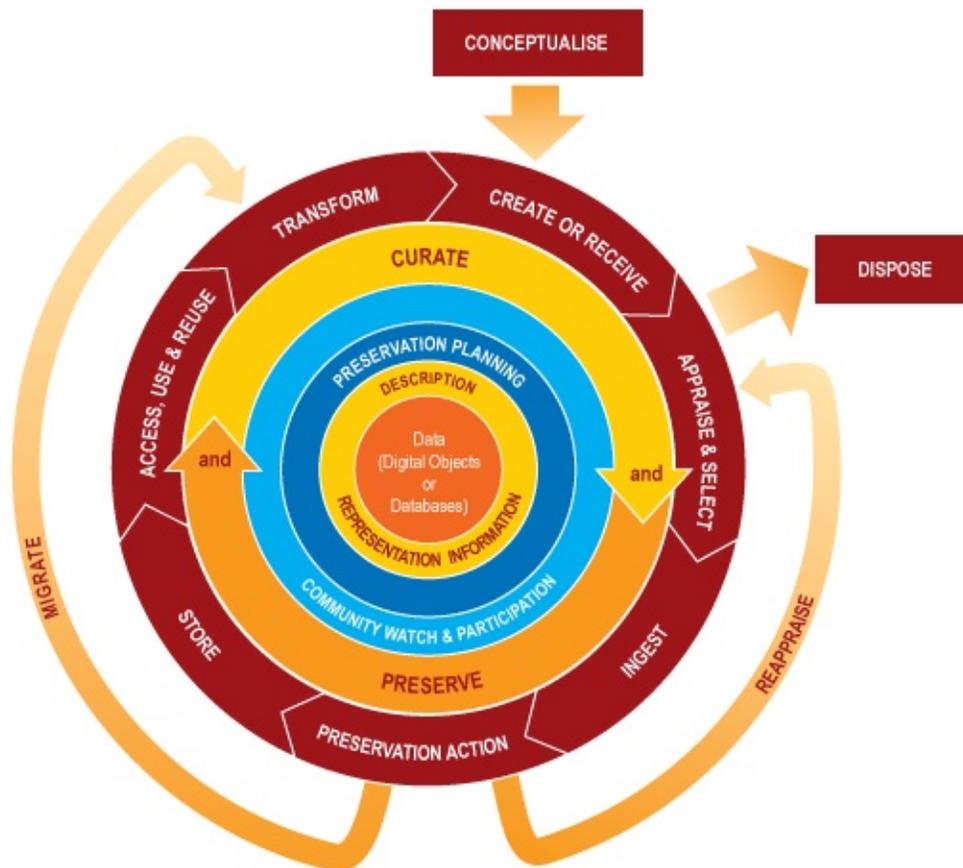
### Table of Contents

Digital Curation Centre (DCC) Lifecycle Model  
Ellyn Montgomery, USGS, Data Lifecycle Diagram  
FGDC Stages of the Geospatial Data Lifecycle pursuant to OMB Circular A-16  
University of Oxford Research Data Management Chart  
NOAA Environmental Data Life Cycle Functions  
Open Archival Information System (OAIS) Framework  
USGS Scientific Information Management Workshop Vocabulary  
Peter Fox Lifecycle Diagrams  
National Science Foundation

## 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 manager 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.

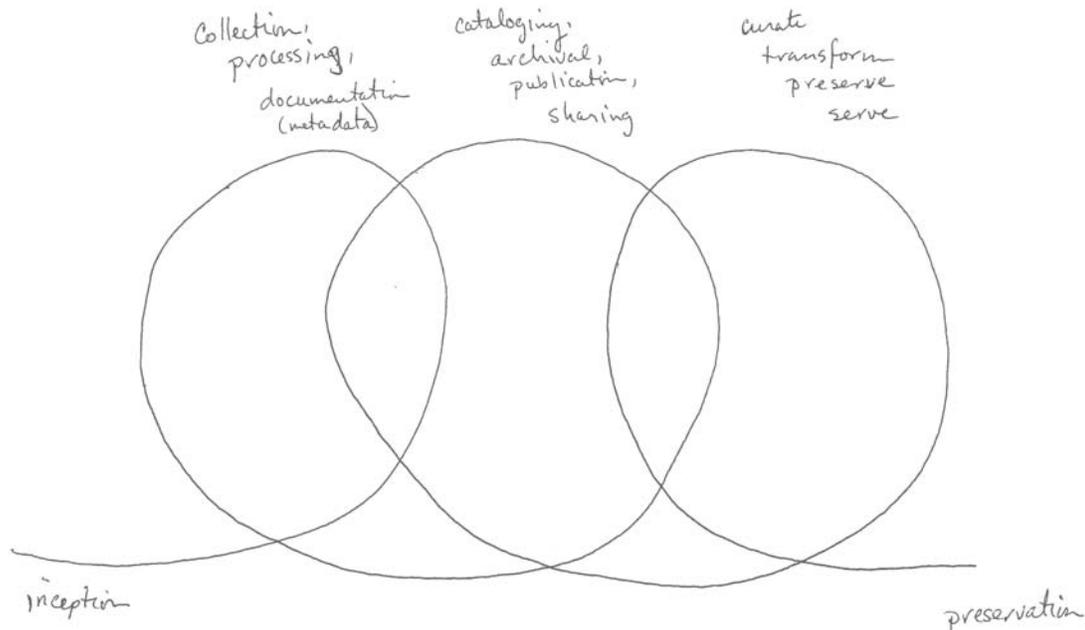
**Migrate**

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.

**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](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

**FGDC STAGES OF THE GEOSPATIAL DATA LIFECYCLE PURSUANT TO  
OMB CIRCULAR A-16, SECTIONS 8(E)(D), 8(E)(F), AND 8(E)(G)**

Found at <http://www.fgdc.gov/policyandplanning/a-16/stages-of-geospatial-data-lifecycle-a16.pdf>

# Stages of the Geospatial Data Lifecycle pursuant to OMB Circular A-16, sections 8(e)(d), 8(e)(f), and 8(e)(g)

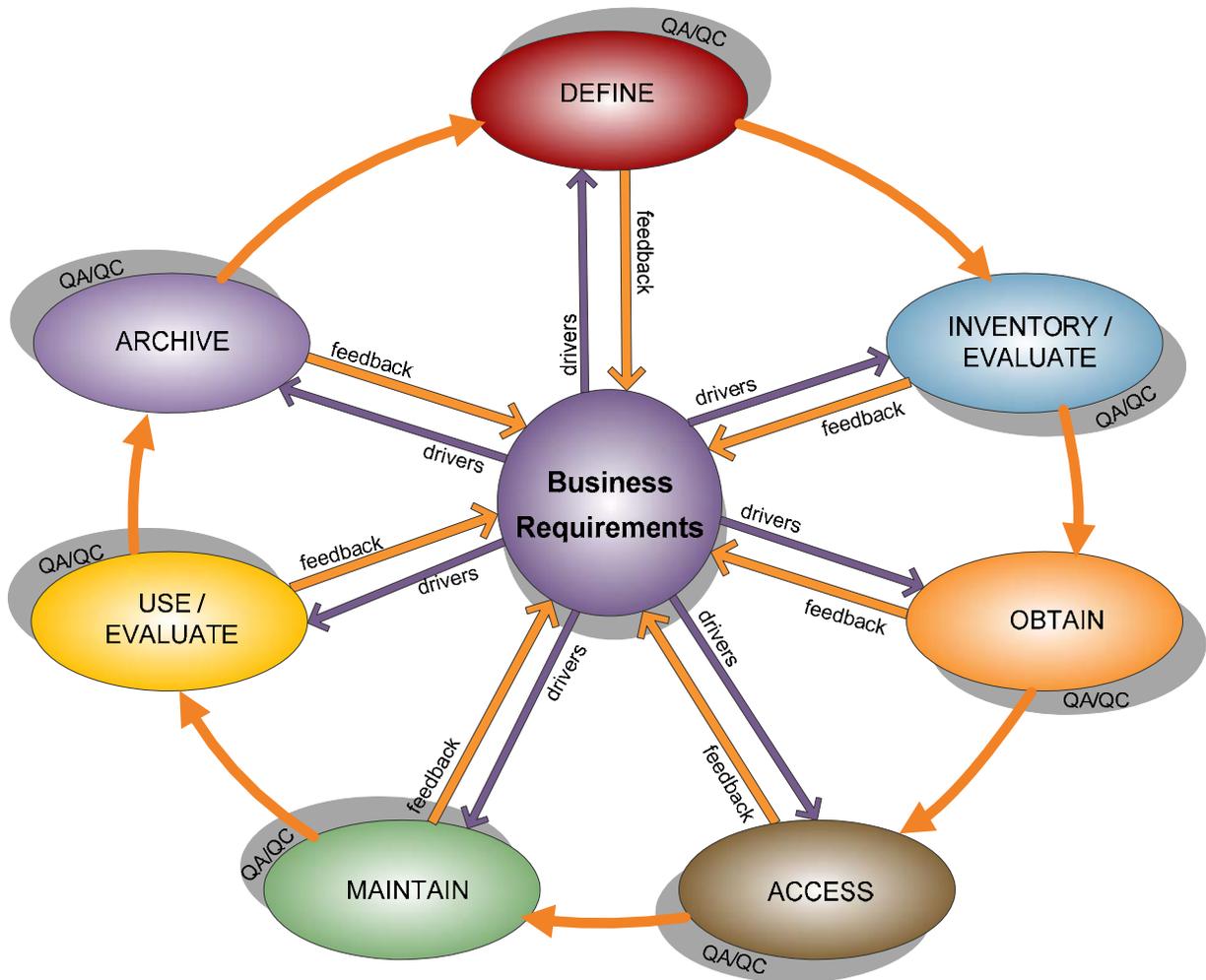
A-16 Supplemental Guidance **Stages of the Geospatial Data Lifecycle** is maintained on the FGDC Website under “Supplemental Guidance Reference Materials.” These reference materials are considered part of the guidance but may be updated as a result of Work Group efforts, FGDC Coordination Group and should be revised and approved by either the FGDC Coordination Group or the FGDC Steering Committee.

This section of the Supplemental Guidance describes the Geospatial Data Lifecycle stages that agencies should use when developing, managing and reporting on National Geospatial Data Asset (NGDA) Datasets under the auspices of OMB Circular A-16. The matrix establishes a framework of standard terminology and processes for seven Geospatial Data Lifecycle stages. Figure B1 below summarizes the Geospatial Data Lifecycle stages, which are Define, Inventory/Evaluate, Obtain, Access, Maintain, Use/Evaluate and Archive.

The stages associated with the management of the data lifecycle allow stakeholders to assess whether NGDA data production activities meet business requirements and utilize best practices that enable shared or common services. The Geospatial Data Lifecycle is not intended to be rigidly sequential or linear. The quality assurance and (or) quality control (QA/QC) functions for the data should be included at every stage of the Geospatial Data Lifecycle.

Business requirements drive what needs to occur at each stage. This concept is illustrated in Figure B1 by the purple arrows that point from the business requirements circle to each lifecycle stage. The orange arrows that point to the business requirements circle represent the feedback loop that needs to occur at each stage to reassess the business requirements.

Figure B1. The Geospatial Data Lifecycle



The expected outcomes of adopting this framework include:

- 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
- Improved strategies for completing and maintaining nationally significant themes and datasets associated with OMB Circular A-16 to enhance services to citizens.

**Table B1. Stages of the Geospatial Data Lifecycle**

Stage	Data Lifecycle Stage Details
<b>Stage 1</b>	<b>Define</b>
Definition	Characterization of data requirements based upon business-driven user needs
User requirements	<p>The following steps occur when defining data needs:</p> <ul style="list-style-type: none"> <li>• Define user requirements               <ul style="list-style-type: none"> <li>○ Identify user or business needs                   <ul style="list-style-type: none"> <li>▪ Review organizational documentation (many organizations publish information on business requirements)</li> <li>▪ Evaluate internal and external needs</li> <li>▪ Identify available funding and resources</li> <li>▪ Review current collaboration efforts</li> <li>▪ Review and evaluate existing knowledge base</li> <li>▪ Identify business data quality requirements (including both locational and positional accuracy and quality measures for the attribute information associated with the geospatial features)</li> </ul> </li> <li>○ Conduct formal requirements collection process, including business process modeling or workflow modeling</li> <li>○ Identify agency or program requirements                   <ul style="list-style-type: none"> <li>▪ Determine the levels of government involved</li> <li>▪ Identify sources for requirements                       <ul style="list-style-type: none"> <li>• Examples: FGDC standards, existing program and project requirements, surveys, workshops and conferences, town hall meetings, annual meetings, discussions with subject matter experts</li> </ul> </li> <li>▪ Determine if requirement is annual or recurring versus one-time (for example, a special focus project)</li> <li>▪ On a case-by-case basis, determine the user requirements for the following:                       <ul style="list-style-type: none"> <li>• Resources (human, capital, financial, and technical)</li> <li>• Resource availability</li> <li>• Project schedules</li> </ul> </li> </ul> </li> <li>○ Evaluate available funds or sources of funding for supporting identifying business needs and agency or program requirements</li> <li>○ Document business needs and agency or program requirements</li> </ul> </li> <li>• Define high-level business process work flows               <ul style="list-style-type: none"> <li>○ Document business process work flows</li> <li>○ Define update cycle for documenting business process work flows</li> <li>○ Define roles and responsibilities for documenting and maintaining business process workflows and definitions thereof</li> </ul> </li> <li>• Define information workflows and data requirements for business processes at a high level               <ul style="list-style-type: none"> <li>○ Identify sources for information workflows and data requirements</li> <li>○ Capture and document information workflows and data requirements</li> <li>○ Develop a data dictionary and/or interface dictionary</li> </ul> </li> </ul>

Stage	Data Lifecycle Stage Details
Stage 1	Define
	<ul style="list-style-type: none"> <li>○ Develop an entity relationship diagram</li> <li>• Define data needs, conduct formal data modeling and document data needs assessment <ul style="list-style-type: none"> <li>○ Determine appropriate feature types</li> <li>○ Identify attributes needed</li> <li>○ Determine data schema</li> <li>○ Identify accuracy levels for individual attributes and for the dataset as a whole</li> <li>○ Determine status of data timeliness (that is, how current are the data)</li> <li>○ Determine data format needs</li> <li>○ Determine appropriate resolution or scale of data</li> <li>○ Identify any seasonal concerns with data collection, acquisition, or production</li> <li>○ Determine data specifications (that is, raster or vector, infrared, or black and white)</li> <li>○ Sample method 1: business process reengineering to determine data needs <ul style="list-style-type: none"> <li>▪ Complete a series of process modeling sessions to document the business processes of the organization</li> <li>▪ Determine what information flows are generated as input for data needs</li> <li>▪ Determine what information flows are generated as output data</li> <li>▪ Create a logical data model of those information flows to describe data needs</li> <li>▪ Develop use cases for the business areas identified</li> </ul> </li> <li>○ Sample method 2: GIS analysis <ul style="list-style-type: none"> <li>▪ Perform use case analyses to determine actual data requirements</li> <li>▪ Define data in geospatial terms of features and attributes</li> </ul> </li> </ul> </li> <li>• Define data standards<sup>1</sup> <ul style="list-style-type: none"> <li>○ Data standards refer to the name, definition, presentation, and business rules governing datasets; they are based on the known data requirements, are set by the stakeholders who need the data and can cross organizational boundaries</li> <li>○ Define data standards for production of data layers using logical, physical or other data models</li> <li>○ Data standards should be defined before data are collected. Advantages to having a standard before data collection include the following <ul style="list-style-type: none"> <li>▪ Ensuring users have the right data to accomplish the task</li> <li>▪ Knowing what the data are once they are collected</li> <li>▪ Having the ability to reproduce the data and collect additional matching data at a later date that can be used in analysis</li> </ul> </li> </ul> </li> </ul>

<sup>1</sup> For more information about data standards, please refer to ISO/IEC 15288:2008, 2008, Systems and software engineering -- System lifecycle processes (Edition 2), International Organization for Standardization (<http://www.iso.org>) available from: [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=43564](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43564)

Stage	Data Lifecycle Stage Details
Stage 1	Define
	<ul style="list-style-type: none"> <li>○ Document and publish data standards</li> <li>● Evaluate past data uses for programs and projects and verify that past data products met user business requirements <ul style="list-style-type: none"> <li>○ Example: project driven reports that take place over a 2- to 3-year span</li> <li>○ Identify geospatial data that can fill those needs</li> <li>○ Consider data schemas that will allow for change analysis over time; standards based on logical data models can be flexible to adapt to changing physical environments</li> <li>○ Address potential integration issues (that is, quality, consistency, and compatibility) before data are collected, including: standards, geo id and datum, positional accuracy, classification system (scheme), <i>in situ</i> sampling logic, census enumeration logic, metadata collection, address matching, and privacy issues</li> </ul> </li> <li>● Determine if data required contain any information governed by the Privacy Act of 1974 <ul style="list-style-type: none"> <li>○ Conduct and document a Privacy Act assessment</li> </ul> </li> <li>● Determine if data required will be of a sensitive nature and require any special handling <ul style="list-style-type: none"> <li>○ Sensitive data may include: <ul style="list-style-type: none"> <li>▪ Data that can be collected but that should not be shared with the public [for example, cultural resources data or personally identifiable information (PII)]</li> <li>▪ Data that must be generalized before release</li> <li>▪ Data that may not be sensitive by itself but which may become sensitive when combined with other datasets</li> <li>▪ Records requirements which may determine whether data will become a dataset of record</li> </ul> </li> <li>○ If data are of a sensitive nature, conduct and document sensitivity evaluation</li> <li>○ Define and document roles and responsibilities associated with a re-evaluation</li> </ul> </li> <li>● Define and develop quality assurance and (or) quality control (QA/QC) measures <ul style="list-style-type: none"> <li>○ Develop QA/QC measures to match the business requirements for the dataset</li> <li>○ Define processes and measures on datasets that will ensure QA/QC including certification processes</li> </ul> </li> <li>● Identify dataset ties to the Federal Enterprise Architecture (FEA) Business Reference Models (BRMs) and Data Reference Models (DRMs)</li> </ul>

Stage	Data Lifecycle Stage Details
<b>Stage 2</b>	<b>Inventory/Evaluate</b>
Definition	The creation and publication of a detailed list of data assets and data gaps (both internal and external) as they relate to business-driven user needs
	<ul style="list-style-type: none"> <li>• Identify available data sources both internal and external to the organization that may meet the defined data needs (Research internal assets first; research external assets second) <ul style="list-style-type: none"> <li>○ Catalog data that may meet defined standards</li> <li>○ Inventory list should be regularly maintained</li> </ul> </li> <li>• Take the following steps to search for defined requirements and create an inventory matrix <ul style="list-style-type: none"> <li>○ Define the area(s) of interest (AOI) for the data</li> <li>○ Locate viable resources for required data (including internal and external sources)</li> <li>○ Identify internal resources that may need conversion</li> <li>○ Review dataset descriptions, metadata records, attributes, domains, etc.</li> <li>○ Address accuracy issues and tiers of accuracy</li> <li>○ Review date ranges and currency of the dataset(s) and identify limitations</li> <li>○ Include restrictions, licensing, and copyright status of obtained datasets documented in a matrix</li> <li>○ Identify closely matched datasets</li> </ul> </li> <li>• The results of the inventory will be a matrix of available geospatial data, which helps users to <ul style="list-style-type: none"> <li>○ Exhaust all possible means to identify existing data <ul style="list-style-type: none"> <li>▪ Examples: <ul style="list-style-type: none"> <li>• Preliminary list of available data and their characteristics (GIS Inventory System, GOS, Web searches)</li> <li>• <a href="http://www.statelocalgov.net">www.statelocalgov.net</a> (site for finding data inventories)</li> <li>• Data centers provide points of contacts for sources and base maps</li> </ul> </li> </ul> </li> <li>○ Identify attributes and domains <ul style="list-style-type: none"> <li>▪ Determine if data meet the identified business requirements</li> <li>▪ Determine if data are missing requirements</li> </ul> </li> <li>○ Assess the suitability of data for meeting the business requirements identified and perform a QA/QC evaluation <ul style="list-style-type: none"> <li>▪ Reject metadata as unsuitable or incomplete</li> <li>▪ Request additional metadata information</li> </ul> </li> <li>○ Identify suitability of dataset to meet business needs <ul style="list-style-type: none"> <li>▪ Evaluate free datasets</li> <li>▪ Obtain datasets</li> <li>▪ Perform a QA/QC evaluation</li> <li>▪ Document whether not data are acceptable and, if not acceptable, find alternative solutions</li> </ul> </li> <li>○ Evaluate datasets for price and develop a financial plan</li> </ul> </li> </ul>

Stage	Data Lifecycle Stage Details
<b>Stage 2</b>	<b>Inventory/Evaluate</b>
	<ul style="list-style-type: none"> <li>▪ Review datasets carefully against business requirements</li> <li>▪ Check with other sources, previous users, technical expert specifications, and so forth, to ensure that data will meet the requirements</li> <li>▪ Example: cost of imagery acquisition or cost of LiDAR acquisition</li> <li>• Determine the limitations of what is found against the requirements to ensure that the data can be integrated for analysis <ul style="list-style-type: none"> <li>○ Evaluate data to ensure it can be integrated</li> <li>○ Evaluate whether data should be made available for public distribution or limited distribution (for example, contractual data, classified data, public domain data, and so forth)</li> </ul> </li> <li>• Assess conversion, transformation, or modification of closely matched datasets to meet requirements</li> <li>• Recommend specialized procedures or protocols to acquire datasets that meet the specified requirements <ul style="list-style-type: none"> <li>○ Develop feedback process(es)/loop(s) for refining data to meet data needs</li> </ul> </li> <li>• Perform a metadata evaluation <ul style="list-style-type: none"> <li>○ Identify or define the metadata standard for use</li> <li>○ Document the metadata evaluation</li> </ul> </li> </ul>

Stage	Data Lifecycle Stage Details
Stage 3	Obtain
Definition	The collection, purchase, conversion, transformation, sharing, exchange, or creation of geospatial data that were selected to meet the business needs is identified
	<ul style="list-style-type: none"> <li>• Review method(s) used to obtain, create, change or acquire datasets necessary for meeting program or project needs <ul style="list-style-type: none"> <li>○ Include methods used to obtain data in the feedback loop</li> </ul> </li> <li>• Follow the process below to obtain data: <ul style="list-style-type: none"> <li>○ Develop and document a plan to obtain data and review results with the program or project documents that defined the requirements <ul style="list-style-type: none"> <li>▪ Complete writing of dataset specifications and statements of work (SOWs)</li> <li>▪ Identify and review similar contracts and SOWs from the immediate agency</li> <li>▪ Utilize agency expertise to review contract specifications</li> <li>▪ Ensure that geospatial requirements boilerplate language is included in planning documents</li> <li>▪ Establish a plan for reviewing results of the process used to obtain data</li> </ul> </li> <li>○ Identify the appropriate contracting officer's technical representative (COTR); that is, one with geospatial capabilities that allow for appropriate QA/QC of the dataset</li> <li>○ Ensure that metadata meet the requirements of the organization; that is, make sure that all metadata fields are completed, not just the mandatory metadata elements (sections 1 and 7) <ul style="list-style-type: none"> <li>▪ Develop a metadata template</li> <li>▪ Evaluate/review metadata <ul style="list-style-type: none"> <li>• Reject metadata that are unsuitable or incomplete</li> <li>• Request additional metadata information as needed</li> </ul> </li> <li>▪ Where appropriate, expand beyond FGDC metadata standards</li> </ul> </li> <li>○ Use inventory lists to identify agencies or entities that have needed data to see if you can establish a partnership to leverage acquisition costs and activities <ul style="list-style-type: none"> <li>▪ Document return-on-investment of partnerships</li> </ul> </li> <li>○ Review agreements that may limit or enhance the sharing and exchange of information <ul style="list-style-type: none"> <li>▪ Define and document roles and responsibilities for reviewing agreements</li> <li>▪ Document review of agreements</li> <li>▪ Agreements may include the following: Memoranda of understanding and service-level agreements (SLAs) <ul style="list-style-type: none"> <li>• Interagency agreements</li> <li>• Military interagency purchase requests</li> <li>• Multiagency purchase requests</li> <li>• Cost-sharing activities</li> <li>• Collaboration activities</li> <li>• Free exchange sites</li> </ul> </li> </ul> </li> </ul> </li> </ul>

Stage	Data Lifecycle Stage Details
Stage 3	Obtain
	<ul style="list-style-type: none"> <li>• In-kind services</li> <li>• Adhere to existing data standards<sup>2</sup></li> <li>• Define and document data collection specifications and update/refresh cycle</li> <li>• Obtain datasets necessary to meet program or project needs in one of the following four ways:             <ol style="list-style-type: none"> <li>1) <b>Purchase</b> <ul style="list-style-type: none"> <li>○ Identify outside sources or other government agencies that currently have (or have access to) the needed data</li> <li>○ Acquire and assess the suitability of data (using a decision tree) and perform QA/QC evaluation                   <ul style="list-style-type: none"> <li>▪ Identify required data items that are essential for acquisition of the dataset</li> </ul> </li> <li>○ Use QA/QC measures established during the Define stage to perform measurements of the data to determine suitability                   <ul style="list-style-type: none"> <li>▪ Determine the data format; geospatial data will be in an agreed upon format (for example, shape files, geodatabases, and so forth)</li> <li>▪ Complete the metadata with respect to contract specifications</li> <li>▪ Check to see that the required metadata sections are populated; the content must be validated by knowledgeable staff to assure compliance                       <ul style="list-style-type: none"> <li>• Complete mandatory metadata sections 1 and 7 (recommended is 1 through 7)</li> <li>• Complete intermediate requirements for the specified sections</li> <li>• Obtain full FGDC compliance in all sections</li> </ul> </li> </ul> </li> <li>○ Establish database schema validation, including the following:                   <ul style="list-style-type: none"> <li>▪ Dataset name and organization</li> <li>▪ Attributes and relationships</li> <li>▪ Domains, topology, and so forth</li> <li>▪ Ability to automate based on database schema required and organizational business requirements</li> </ul> </li> <li>○ Determine projection (spatial reference) and ensure that the coordinate system is fully defined</li> <li>○ Ensure that the metadata meet the requirements of the organization; that is, make sure that all metadata fields are completed</li> <li>○ Consider the level at which the data are certified                   <ul style="list-style-type: none"> <li>▪ Verify the certification level of the data</li> <li>▪ Evaluate the data against the information required by the Information Quality Act</li> <li>▪ If data are not certified, establish a certification plan</li> </ul> </li> <li>○ Review licensing agreements and any restrictions on release of the information (the Freedom of Information Act may come into play here)</li> </ul></li></ol> </li> </ul>

<sup>2</sup> For more information about data standards, please refer to ISO/IEC 15288:2008, 2008, Systems and software engineering -- System lifecycle processes (Edition 2), International Organization for Standardization (<http://www.iso.org>). Available from: [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=43564](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43564)

Stage	Data Lifecycle Stage Details
Stage 3	Obtain
	<ul style="list-style-type: none"> <li>▪ Examples of purchased data: U.S. Geological Survey (USGS) Center for Earth Resources Observation and Science (EROS) satellite imagery, aerial photography</li> <li>▪ Vehicles of Procurement: request for proposal (RFP), License, request for information (RFI), General Services Administration (GSA) Schedule, grants, SLAs, best practices</li> <li>▪ Review agreements to share the data (see Geospatial LoB grants and contracts language)</li> </ul> <p><b>2) Modify, convert, or transform legacy data</b></p> <p><u>Modify</u></p> <ul style="list-style-type: none"> <li>○ Identify missing requirements and correct the following: <ul style="list-style-type: none"> <li>▪ Attributes – add necessary attributes</li> <li>▪ Domain – populate or modify</li> <li>▪ Status of data timeliness – if not current, update as necessary</li> <li>▪ Completeness – fill in or add additional feature types, as needed</li> </ul> </li> <li>○ Review the quality and suitability of data for use by geospatially knowledgeable staff and program or project managers <ul style="list-style-type: none"> <li>▪ Use QA/QC measures established during the Define stage to perform measurements of data to determine suitability</li> <li>▪ Establish the data format – geospatial data will be in an agreed-upon format (for example, shape files, geodatabases, and so forth)</li> <li>▪ Complete or revise the metadata with respect to contract specifications</li> <li>▪ Check that the required sections are populated; have the content validated by knowledgeable staff to assure compliance <ul style="list-style-type: none"> <li>• Complete mandatory metadata sections 1 and 7 (recommended is 1 through 7)</li> <li>• Complete intermediate requirements for the specified sections</li> <li>• Obtain full FGDC compliance in all sections</li> </ul> </li> </ul> </li> <li>○ Establish database schema validation including the following: <ul style="list-style-type: none"> <li>▪ Dataset name and organization</li> <li>▪ Attributes and relationships</li> <li>▪ Domains, topology, and so forth</li> <li>▪ Ability to automate based on database schema required and organizational business requirements</li> </ul> </li> <li>○ Determine projection (spatial reference) and ensure that the coordinate system is fully defined</li> </ul> <p><u>Convert/Transform</u></p> <ul style="list-style-type: none"> <li>○ Review format, data storage method, and data field information and determine what items need conversion and (or) transformation</li> </ul>

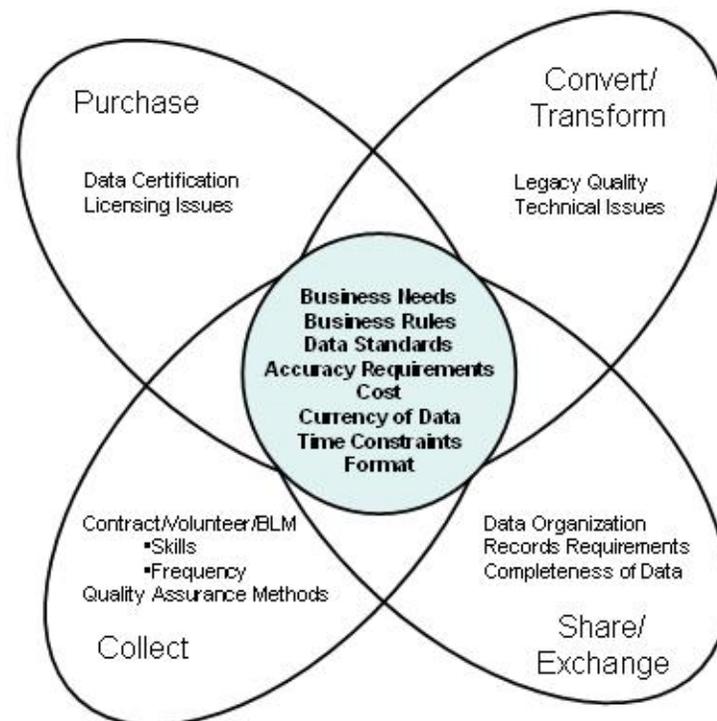
Stage	Data Lifecycle Stage Details
Stage 3	Obtain
	<ul style="list-style-type: none"> <li>○ Evaluate the extract, transform, load (ETL) hardware or software<sup>3</sup></li> <li>○ Evaluate the conversion and (or) transformation tools for compatibility with data formats</li> <li>○ After conversion/transformation, determine if legacy data will meet user needs</li> <li>○ Review the quality and suitability of the data for use by geospatially knowledgeable staff and program or project manager <ul style="list-style-type: none"> <li>▪ Use QA/QC measures established during the Define stage to perform measurements of the data to determine suitability</li> <li>▪ Establish the data format – geospatial data will be in an agreed upon format (for example, shape files, geodatabases, and so forth)</li> <li>▪ Complete or revise the metadata with respect to contract specifications</li> <li>▪ Check to see that the required sections are populated; have the content validated by knowledgeable staff to assure compliance <ul style="list-style-type: none"> <li>• Complete mandatory metadata sections 1 and 7 (recommended is 1 through 7)</li> <li>• Complete intermediate requirements for the specified sections</li> <li>• Obtain full FGDC compliance in all sections</li> </ul> </li> </ul> </li> <li>○ Establish database schema validation, including the following: <ul style="list-style-type: none"> <li>▪ Dataset name and organization</li> <li>▪ Attributes and relationships</li> <li>▪ Domains, topology, and so forth</li> <li>▪ Ability to automate based on the database schema required and organizational business requirements</li> </ul> </li> <li>○ Determine projection (spatial reference) – coordinate system is fully defined</li> <li><b>3) Share/Exchange with others</b></li> <li>○ Identify agencies or entities that have needed data</li> <li>○ Review agreements that may limit or enhance the sharing and exchange of information (see list at beginning of Stage 3)</li> <li>○ Consider how the data are organized and how the data relate to user business needs</li> <li>○ Identify the data sharing method as distinguished by the OMB <ul style="list-style-type: none"> <li>▪ Data exchanges <ul style="list-style-type: none"> <li>• Determine permissions involved for one or more persons to input information into a common data store or information system</li> <li>• Update data in real-time as the entries are made</li> <li>• Access permissions made to the individual per agreement between and among organizations</li> </ul> </li> <li>▪ Data-sharing</li> </ul> </li> </ul>

<sup>3</sup> ETL hardware/software assists in the process of defining and applying algorithms to change data from one form or domain value set to another domain value set in the target architecture

Stage	Data Lifecycle Stage Details
Stage 3	Obtain
	<ul style="list-style-type: none"> <li>• Establish permission for each party to retain its own database or application and for the other party to be permitted to read the information</li> <li>• Give access rights to individuals from the other organization but on a “read only” capability</li> <li>• Allow for agencies to do one-time-only data collection yet use information across the agencies</li> <li>○ Review the quality and suitability of the data for use by geospatially knowledgeable staff and program or project managers <ul style="list-style-type: none"> <li>▪ Use QA/QC measures established during the Define stage to perform measurements of the data to determine suitability</li> <li>▪ Establish the data format – geospatial data will be in an agreed upon format (for example, shape files, geodatabases, etc.)</li> <li>▪ Complete or revise the metadata with respect to contract specifications</li> <li>▪ Check to see that the required sections have are populated; have the content validated by knowledgeable staff to assure compliance <ul style="list-style-type: none"> <li>• Complete mandatory metadata sections 1 and 7 (recommended is 1 through 7)</li> <li>• Complete intermediate requirements for the specific sections</li> <li>• Obtain full FGDC compliance in all sections</li> </ul> </li> </ul> </li> <li>○ Establish database schema validation, including the following: <ul style="list-style-type: none"> <li>▪ Dataset name and organization</li> <li>▪ Attributes and relationships</li> <li>▪ Domains, topology, etc.</li> <li>▪ Ability to automate based upon the database schema required and organizational business requirements</li> </ul> </li> <li>○ Determine projection (spatial reference) and ensure that the coordinate system is fully defined <ul style="list-style-type: none"> <li>▪ Reject the dataset</li> <li>▪ Accept the dataset</li> </ul> </li> <li><b>4) Create</b> <ul style="list-style-type: none"> <li>○ Using the inventory matrix from the Inventory/Evaluate stage, determine which datasets meet the needs of the program or project</li> <li>○ Create new geospatial datasets from analog maps, printed reports, archived resources, etc. <ul style="list-style-type: none"> <li>▪ Evaluate requirements for collection of data <ul style="list-style-type: none"> <li>• Scheduled timeframe for collection of data</li> <li>• Certifications, user needs grants and contracts language, etc.</li> <li>• Frequency of collection (that is, recurring versus special use)</li> </ul> </li> <li>▪ Evaluate internal and external funding resources for the collection of data</li> <li>▪ Enter data – digitize, key entry, scan, and etc.</li> </ul> </li> </ul> </li> </ul>

Stage	Data Lifecycle Stage Details
Stage 3	Obtain
	<ul style="list-style-type: none"> <li>▪ Complete or revise metadata (recommended sections 1 through 7)</li> <li>○ Review and acceptance of data by geospatially knowledgeable staff and by program or project manager</li> </ul>

Figure B2: Considerations for acquisition of datasets



Source: Bureau of Land Management data stewardship training

Stage	Data Lifecycle Stage Details
<b>Stage 4</b>	<b>Access</b>
Definition	Making data produced known and retrievable to the community through documentation and discovery mechanisms so the users can meet their business requirements
	<ul style="list-style-type: none"> <li>• Review plans for community at large user access <ul style="list-style-type: none"> <li>○ Develop a User Access Plan for dissemination/access that indicates who should have access to your data and how data are disseminated</li> <li>○ Document roles and responsibilities around determining where data are offered</li> </ul> </li> <li>• Review structure for access, user groups and individuals <ul style="list-style-type: none"> <li>○ Develop a high level CRUD (Create, Read, Use, Delete) type matrix showing who should have access to create, read, use, or delete data</li> <li>○ Perform access management review</li> <li>○ Create and publish a records schedule</li> </ul> </li> </ul> <p><b>1) Publish – Publication Plan</b></p> <ul style="list-style-type: none"> <li>• Determine and document where to publish the data or offering of services <ul style="list-style-type: none"> <li>○ Consider the user community and determine if the data should be made available in hard copy, harvested by the GOS portal or other portals, provided to a clearinghouse, and so forth (for example, GOS existing, planned, and new data)</li> <li>○ Review user needs and business requirements</li> <li>○ Review Geospatial LoB grants and contracts language for publishing restrictions or requirements</li> <li>○ Review regulations for requirements <ul style="list-style-type: none"> <li>▪ Example: Executive Order No. 12906, section 3(c) (April 11, 1994) mandates that each agency must have procedures to make geospatial information available to the public</li> </ul> </li> <li>○ Include financial commitments as part of the publication plan</li> </ul> </li> <li>• Determine the format in which the data will be published (for example, shape files, E00 format, data streaming, Extensible Markup Language (XML), Generalized Markup Language (GML), etc.) <ul style="list-style-type: none"> <li>○ Include formatting requirements as part of our QA/QC plan and feedback loop</li> </ul> </li> <li>• Comply with any formatting requirements, including the following: <ul style="list-style-type: none"> <li>○ Metadata standards (recommended minimum 1 through 7)</li> <li>○ Attribution for legacy data migration that links back to source</li> <li>○ Agency requirements</li> <li>○ FGDC requirements</li> <li>○ Federal guidelines</li> <li>○ Executive Order 12906 requirements</li> <li>○ OMB Circular A-16 requirements</li> <li>○ GOS guidance</li> </ul> </li> <li>• Determine where partnerships might exist to meet interagency and other stakeholder needs</li> <li>• Comply with electronic data management requirements for the following regulations and any others that may apply:</li> </ul>

Stage	Data Lifecycle Stage Details
Stage 4	Access
	<ul style="list-style-type: none"> <li>○ Determine what effects the Freedom of Information Act (FOIA)<sup>4</sup> and the Privacy Act<sup>5</sup> may have on publishing the dataset</li> <li>○ Perform a privacy impact assessment (PIA)<sup>6</sup></li> <li>○ Perform Federal Information Security Management Act (FISMA) review.<sup>7</sup> All data must have a security review and be categorized (that is, establish security controls that prevent harm or protect integrity and confidentiality)</li> <li>○ Determine type of records that may be produced from data according to the National Archives and Records Administration (NARA)<sup>8</sup></li> </ul> <p><b>2) Disseminate –Dissemination Plan</b></p> <ul style="list-style-type: none"> <li>● Consider data organization for dissemination</li> <li>● Perform records management review (create and publish a records schedule)</li> <li>● Evaluate data steward data completeness <ul style="list-style-type: none"> <li>○ Determine if agencies have a dissemination plan for internal and external stakeholders</li> <li>○ Determine if agencies have a format for dissemination</li> <li>○ Determine if data completeness changes with stakeholders</li> <li>○ Review Web feature services, specific requests, and agency restrictions (for example, if an agency does not permit FTP sites or has security restrictions)</li> <li>○ Determine nonproprietary and proprietary formats (many agencies must disseminate nonproprietary data when they provide proprietary)</li> <li>○ Review Web feature service/Web mapping service and Open Geospatial Consortium (OGC) compliance</li> </ul> </li> <li>● Review dissemination requirements including: <ul style="list-style-type: none"> <li>○ 508 compliance (enables those with visual disabilities to access the information and understand the description of the data, metadata, and so forth)</li> <li>○ Agency requirements</li> <li>○ FGDC requirements</li> </ul> </li> <li>● Determine internal versus external dissemination and access levels including: <ul style="list-style-type: none"> <li>○ Security</li> <li>○ Copyright</li> </ul> </li> </ul>

<sup>4</sup> The Freedom of Information Act (FOIA), Title 5 United States Code, section 552

<sup>5</sup> The Privacy Act of 1974, Pub. L. 93-579, 88 Stat. 1897 (Dec. 31, 1974), codified in part at 5 U.S.C. §552a

<sup>6</sup> Section 208 of the E-Government Act of 2002 (Pub. L. 107-347, 44 U.S.C. Ch 36) requires that OMB issue guidance to agencies on implementing the privacy provisions of the E-Government Act.

<sup>7</sup> The Federal Information Security Management Act of 2002 is the primary legislation governing Federal information security. FISMA was built upon and was an expansion of earlier legislation and added particular emphasis to the management dimension of information security in the Federal Government. FISMA establishes stronger lines of management responsibility for information security and provides for substantial oversight by the legislative branch. It is also called the Electronic Government Act of 2002, Title III of the E-Government Act of 2002, The Federal Information Security Management Act of 2002, FISMA, E-Government Act of 2002, and E-Government Act of 2002.

<sup>8</sup> NARA Regulations are published in the 36 CFR Chapter XII; Electronic Records Management is published under PART 1234 of the 36 CFR Chapter XII

Stage	Data Lifecycle Stage Details
Stage 4	<b>Access</b>
	<ul style="list-style-type: none"> <li>○ Licensing agreements               <ul style="list-style-type: none"> <li>● Document where and how your data are available and accessed</li> <li>● Identify and document user groups</li> <li>● Identify partnerships that may impact dissemination schedule</li> <li>● Evaluate the dissemination plan for your data steward's requirements</li> <li>● Include financial commitments as part of your dissemination plan</li> </ul> </li> </ul>

Stage	Data Lifecycle Stage Details
<b>Stage 5</b>	<b>Maintain</b>
Definition	Ongoing processes and procedures to ensure that the data meet business requirements
	<p><b>1) Maintenance Plan</b></p> <ul style="list-style-type: none"> <li>• Use the steps below to help build a maintenance plan for data and services <ul style="list-style-type: none"> <li>○ Determine hardware and software needed to effectively maintain both data and services</li> <li>○ Determine other resources available to maintain the data <ul style="list-style-type: none"> <li>▪ Establish a strategy to track funding</li> <li>▪ Establish a strategy to track and evaluate partnerships</li> <li>▪ Develop a financial plan for supporting maintenance</li> </ul> </li> <li>○ Determine requirements including: <ul style="list-style-type: none"> <li>▪ Records <ul style="list-style-type: none"> <li>• NARA requirements<sup>9</sup></li> <li>• Business requirements</li> </ul> </li> <li>▪ Retention schedule <ul style="list-style-type: none"> <li>• NARA requirements</li> <li>• Other Government or agency requirements</li> </ul> </li> <li>▪ Snapshots requirements <ul style="list-style-type: none"> <li>• NARA requirements</li> <li>• National Environmental Planning Act (NEPA) requirements</li> </ul> </li> </ul> </li> <li>○ Develop a records plan, retention schedule, and snapshots plan</li> <li>○ Review accuracy of the data, including attributes, to see that they met the business needs of the customer <ul style="list-style-type: none"> <li>▪ Make additions, deletions, and updates to the database in a tightly controlled environment to retain database quality and integrity</li> </ul> </li> <li>○ Create a data QA/QC plan for the dataset <ul style="list-style-type: none"> <li>▪ Establish controls for edits and updates</li> <li>▪ Create review cycles for periodic review of the data against the established QA/QC measures established during the Define stage</li> <li>▪ Complete reviews and measures with a random sampling of the data or by creating a data review cycle before posting the data to the dataset in final form</li> </ul> </li> <li>○ Obtain material from the ground (that is, by way of field data collection) and roll up the information into a maintenance plan</li> <li>○ Establish standards for maintenance that can be updated to keep data in line with user needs</li> <li>○ Establish a standard process for version control</li> </ul> </li> </ul>

<sup>9</sup> 36 CFR Chapter 12, electronic records management

Stage	Data Lifecycle Stage Details
<b>Stage 5</b>	<b>Maintain</b>
	<ul style="list-style-type: none"> <li>▪ If needed, implement manual version control to facilitate dataset maintenance</li> <li>▪ Create static snapshots of the dataset at key critical points (for example, updates, modifications to data structure, format changes, attribute additions or modifications, and so forth)</li> <li>▪ Define and document roles and responsibilities associated with version control</li> <li>○ Develop and document a strategy for updating the process of maintaining the data <ul style="list-style-type: none"> <li>▪ Review responsibilities for maintenance, including the following: <ul style="list-style-type: none"> <li>• Theme Lead roles</li> <li>• Data steward roles</li> <li>• MOUs between agencies and partners (outline an agreed-upon process)</li> <li>• National and State stewardship plans</li> </ul> </li> <li>▪ Assign specific maintenance responsibilities</li> <li>▪ Update financial plans as part of this strategy</li> <li>▪ Notify stewards and end-users of updates</li> </ul> </li> <li>○ Define and develop performance expectations for SLAs as part of service-oriented architecture (SOA)</li> <li>○ Establish a feedback process to obtain information from users and incorporate feedback into future maintenance plans</li> <li>○ Establish a strategy to retain knowledgeable, experienced and trained staff and make sure that they are using current standards and guidelines <ul style="list-style-type: none"> <li>▪ Document stewardship roles and responsibilities</li> </ul> </li> </ul> <p><b>2) Store and manage</b></p> <ul style="list-style-type: none"> <li>• Develop a plan to store and manage data and services (for example, coverage, map, image services geocoding, tools and applications) including funding plans, system support plans, etc. <ul style="list-style-type: none"> <li>○ Determine the frequency of updates (may vary by dataset or by service) <ul style="list-style-type: none"> <li>▪ Determine if updates are reoccurring or performed on an as needed or as requested basis</li> <li>▪ Establish a point of contact and responsibilities for updates (this is especially important for partnerships)</li> <li>▪ Use metadata to identify data that may need to be updated in a proactive manner</li> </ul> </li> <li>○ Evaluate resources based on the projected frequency of the updates</li> <li>○ Determine resources available, including the following: <ul style="list-style-type: none"> <li>▪ Cost and available funding</li> <li>▪ Staffing available</li> <li>▪ Existing or potential partnerships</li> </ul> </li> <li>○ Determine internal and external storage and management access needs, including the following: <ul style="list-style-type: none"> <li>▪ Level of access (classified versus unclassified)</li> <li>▪ Availability appropriate to the user community</li> </ul> </li> <li>○ Determine container for data storage and management [for example, databases, compact discs (CDs), clearinghouse, etc. ]</li> </ul> </li> </ul>

Stage	Data Lifecycle Stage Details
Stage 5	Maintain
	<ul style="list-style-type: none"> <li>▪ Evaluate internal storage versus external storage</li> <li>▪ Evaluate internal management versus external management</li> <li>▪ Evaluate partnerships in management and storage</li> <li>○ Establish process for version control (manual, automated, hybrid, and so forth)</li> <li>○ Evaluate decision points on data or services that can become key considerations in storage or management</li> <li>○ Evaluate storage and management plan against the FEA Geospatial Profile Version 2.0 or other FEA or segment architecture guidance</li> </ul>

Stage	Data Lifecycle Stage Details
<b>Stage 6</b>	<b>Use/Evaluate</b>
Definition	The ongoing assessment, validation, and potential enhancement of data to meet user needs and business requirements
	<ul style="list-style-type: none"> <li>• Provide education to users including, but not limited to 1) proper metadata documentation; 2) most efficient uses for your data; 3) security requirements, or sensitivity evaluation education; and 4) data effectiveness</li> <li>• Perform the following steps to help ensure efficient and effective use of geospatial assets: <ul style="list-style-type: none"> <li>○ Review requirements to see that use of the asset meets the requirements and intended outcomes <ul style="list-style-type: none"> <li>▪ Include costs for this activity in your financial plan</li> </ul> </li> <li>○ Validate and verify that the data continue to meet business requirements and user needs <ul style="list-style-type: none"> <li>▪ Was information provided in a timely manner?</li> <li>▪ Are attributes meeting user needs?</li> <li>▪ Are domain values appropriate and current?</li> <li>▪ Did the information help the user?</li> <li>▪ Did the user need to modify the information for his or her use?</li> <li>▪ Was the information used to geo-enable other business data?</li> <li>▪ Did or can the information contribute to business process engineering or reengineering within the agency?</li> </ul> </li> <li>○ Evaluate the data as a part of common services across lines of business <ul style="list-style-type: none"> <li>▪ Can data stewards engage in outreach and communication efforts across agencies?</li> <li>▪ Can data be used in other agencies or organizations?</li> <li>▪ Do applications that help others use and re-use the agency data currently exist?</li> </ul> </li> <li>○ Evaluate how resources are being utilized and tracked as the data are used <ul style="list-style-type: none"> <li>▪ Do you measure data usage? How?</li> <li>▪ Do resources used meet expectations?</li> <li>▪ Did the user take advantage of SOA approaches and SLAs?</li> </ul> </li> <li>○ Determine the availability or development of user groups to do the following: <ul style="list-style-type: none"> <li>▪ Share tools</li> <li>▪ Identify and develop services</li> <li>▪ Encourage partnerships and resource sharing</li> <li>▪ Encourage feedback between user and producer of the data</li> </ul> </li> <li>○ Establish a feedback loop (an important element of data stewardship) between end users and producers where end users assist in the evolution of the dataset by communicating how they use the data <ul style="list-style-type: none"> <li>▪ Determine the availability or development of user groups to share tools, identify and develop services, encourage partnerships, etc.</li> </ul> </li> <li>○ Enter analysis information and incorporate it into the feedback loop</li> </ul> </li> </ul>

- Periodically evaluate if the data remain useful
- If data are becoming obsolete, determine why and if action needs to be taken
- Perform QA/QC reviews
  - Evaluate the data against the quality measures established during the first (Define) and second (Inventory/Evaluate) stages of the data lifecycle
  - Review the measures established during the first (Define) and second (Inventory/Evaluate) stages of the data lifecycle if the business requirements have changed
  - Define QA/QC process, roles and responsibilities and schedule
  - Document QA/QC reviews

Stage	Working Stage Details
<b>Stage 7</b>	<b>Archive</b>
Definition	Required retention of data and the data's retirement into long-term storage
	<ul style="list-style-type: none"> <li>▪ Develop a written policy for data disposal in accordance with National Archives and Records Administration (NARA) requirements; NARA is mandated to archive for all agencies <ul style="list-style-type: none"> <li>○ Review requirements for archiving information <ul style="list-style-type: none"> <li>▪ Records management practices established in the Federal Records Act</li> <li>▪ 36 CFR 1200</li> <li>▪ FGDC requirements</li> <li>▪ Agency requirements <ul style="list-style-type: none"> <li>• Document/maintain documented agency archiving policies and requirements</li> </ul> </li> </ul> </li> <li>○ Archive both internally in the organization responsible for the data (offline storage) and externally (snapshot transfers)</li> <li>○ Provide time-stamped versions of copies of data to NARA for long-term preservation in a sustainable format</li> <li>○ Work in conjunction with agency IT and records officers to create and implement an archival plan <ul style="list-style-type: none"> <li>▪ Define and document archiving process, roles and responsibilities and schedule</li> </ul> </li> <li>○ Identify business needs for archiving: <ul style="list-style-type: none"> <li>▪ Isolate datasets to archive; decide whether archived datasets will be composed of the entire dataset each time or subsets of the data <ul style="list-style-type: none"> <li>• Review business requirements and business needs identified in the first stage (Define) of the data lifecycle</li> <li>• Review plans to obtain and maintain data as established in the third (Obtain) and fifth (Maintain) stages of the data lifecycle</li> </ul> </li> <li>▪ Identify available funding and resources</li> <li>▪ Develop a method for transferring permanent data to NARA</li> <li>▪ Dispose of data identified (tied to the plan developed in the Maintain stage)</li> </ul> </li> <li>○ Determine technical requirements including: <ul style="list-style-type: none"> <li>▪ Create data schema to accompany transfer</li> <li>▪ Select sustainable format to store data (tied into original decision about data in the Obtain stage)</li> <li>▪ Establish preservation procedures</li> <li>▪ Sample to make sure that the data are readable</li> </ul> </li> </ul> </li> </ul>

## Differences between the ISO System Lifecycle and the Data Lifecycle

The International Organization for Standardization (ISO) system lifecycle and the Geospatial Data Lifecycle are both necessary. These two lifecycles are not the same or interchangeable, but are complementary. The ISO system lifecycle provides the framework on which to build a specific application that supports the collection, maintenance, and delivery of data to users. It is physically oriented (focused on computer hardware and software) and has a technical lifespan that requires upgrading and eventual replacement because of its dependency on programming and technical specifications. The data lifecycle uses many of the same concepts but focuses specifically on the data, which may exist within a system or be maintained by one or more specific applications. The data and the data lifecycle should be platform and system independent so that the data can evolve or be migrated to a new platform or application as the business requirements change. Unlike an interface, which may change in response to future developments, the data may change in form but not necessarily in content; the data should be multiple-use (that is, the data should be able to be accessed by multiple users, on multiple systems, for multiple purposes).

# **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.



## **NOAA ENVIRONMENTAL DATA LIFE CYCLE FUNCTIONS**

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

# NOAA Environmental Data Life Cycle Functions

## Overarching all aspects of the data life cycle

Governance, Requirements Management, Architecture Management, Security<sup>1</sup>

Developing and maintaining rich metadata to accompany the data<sup>2</sup>

Establishing mechanisms that allow for user requirements and feedback<sup>2</sup>

### Planning of New Observing or Data Management Systems

- Requirements Definition
- Analysis of Alternatives
- System Design
- Integration with observing systems (NOAA, Interagency/ State, International)
- Determining what to archive and associated funding
- Buy/build

### Stewardship Overarches Observing Operations, Archive<sup>2</sup>, Access<sup>2</sup>, Use

-All ongoing, iterative processes that improve:

- a) data & metadata content (includes reprocessing data)<sup>2</sup>
- b) access & user understanding<sup>2</sup>

### Observing Operations

- Actual Observation
- Transmission / Processing / QA
- Integration with other data to create products (e.g., models)
- Dissemination to real-time Subscribers
- Delivery to Archive

### Archive<sup>2,3</sup>

- Ingest (Receipt)
- Archival Storage
- Data management (populating catalogs, registries, metadata)
- Preservation Planning (migration to new technologies)

### Access<sup>2,3</sup>

- Discovery (catalogs, registries, metadata)
- Dissemination to Users (web services, legacy systems, standard formats)

### Use

- Integration w/other information (NOAA, others)
- Assimilation into models
- Product creation
- Make decisions (Policy, emergency, others)
- Scientific discovery
- Feedback to NOAA

<sup>1</sup> Defined by ISO 15489-1: *Information and Documentation - Records Management* (2001)

<sup>2</sup> Defined by NRC: *Environmental Data Management at NOAA - Archiving, Stewardship, and Access* (2007)

<sup>3</sup> Defined by ISO 14721: *Reference Model for an Open Archival Information System (OAIS)* (2002)

## **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

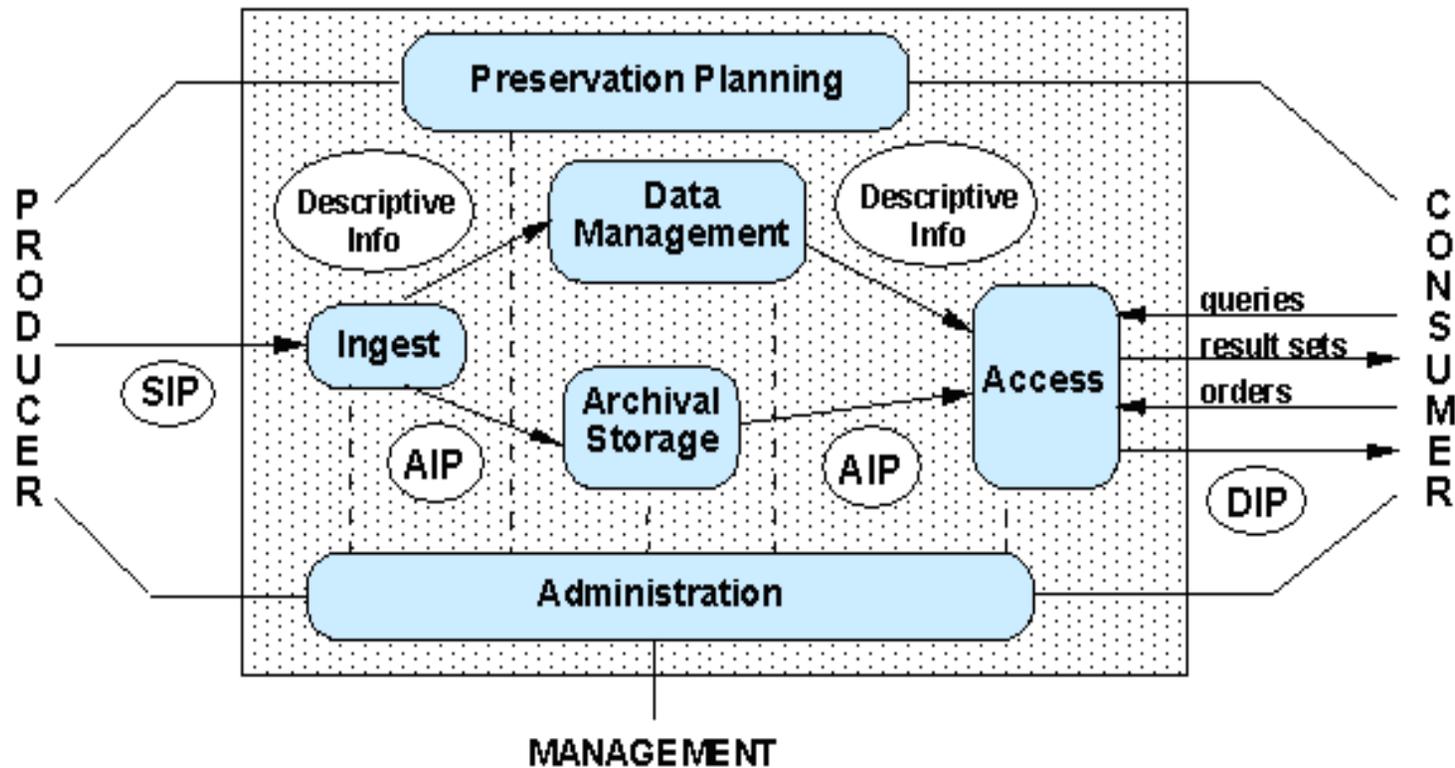
NARA

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 as 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 Digital Preservation Europe (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

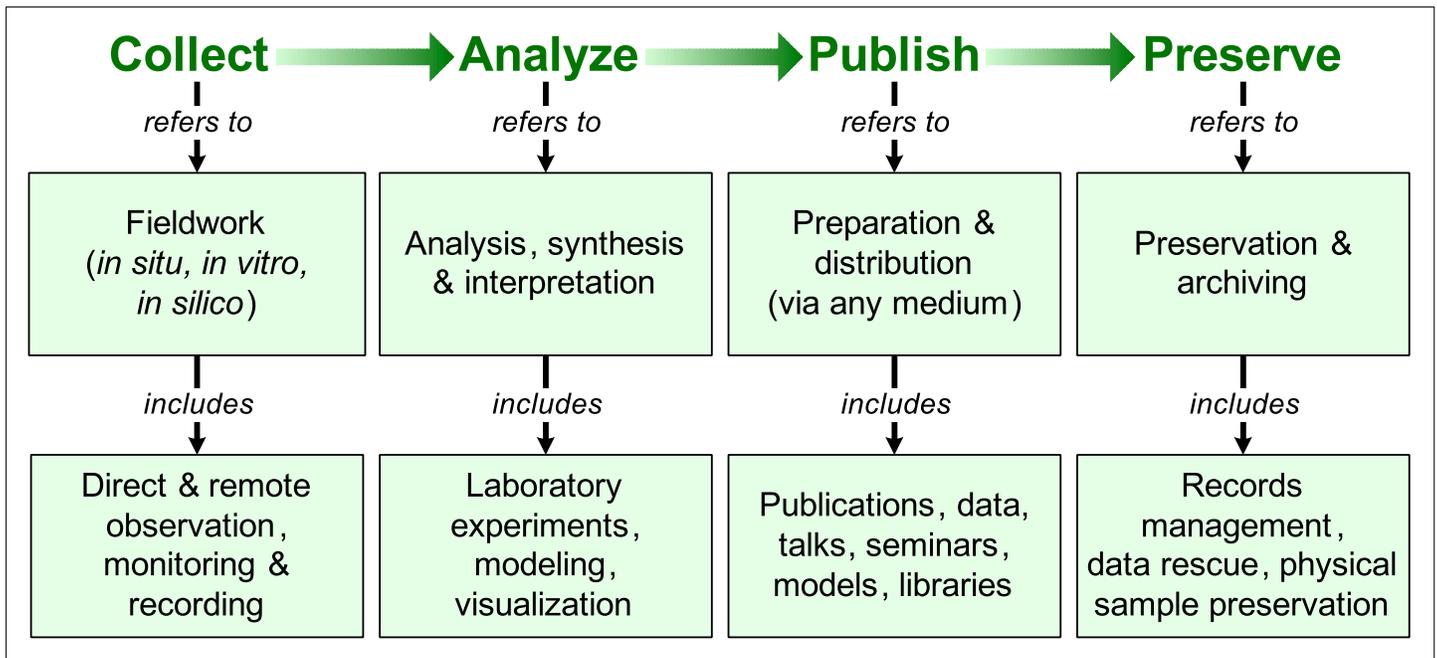
- No integrated “community of practice” for promoting trusted repositories
- CCDSC Red Book and DRAMBORA are not widely known in the scientific, IT, and data management communities
- They do not address design of data systems and data management, e.g. systems planning, funding, system requirements
-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

## **USGS Scientific Information Management Workshop Vocabulary**

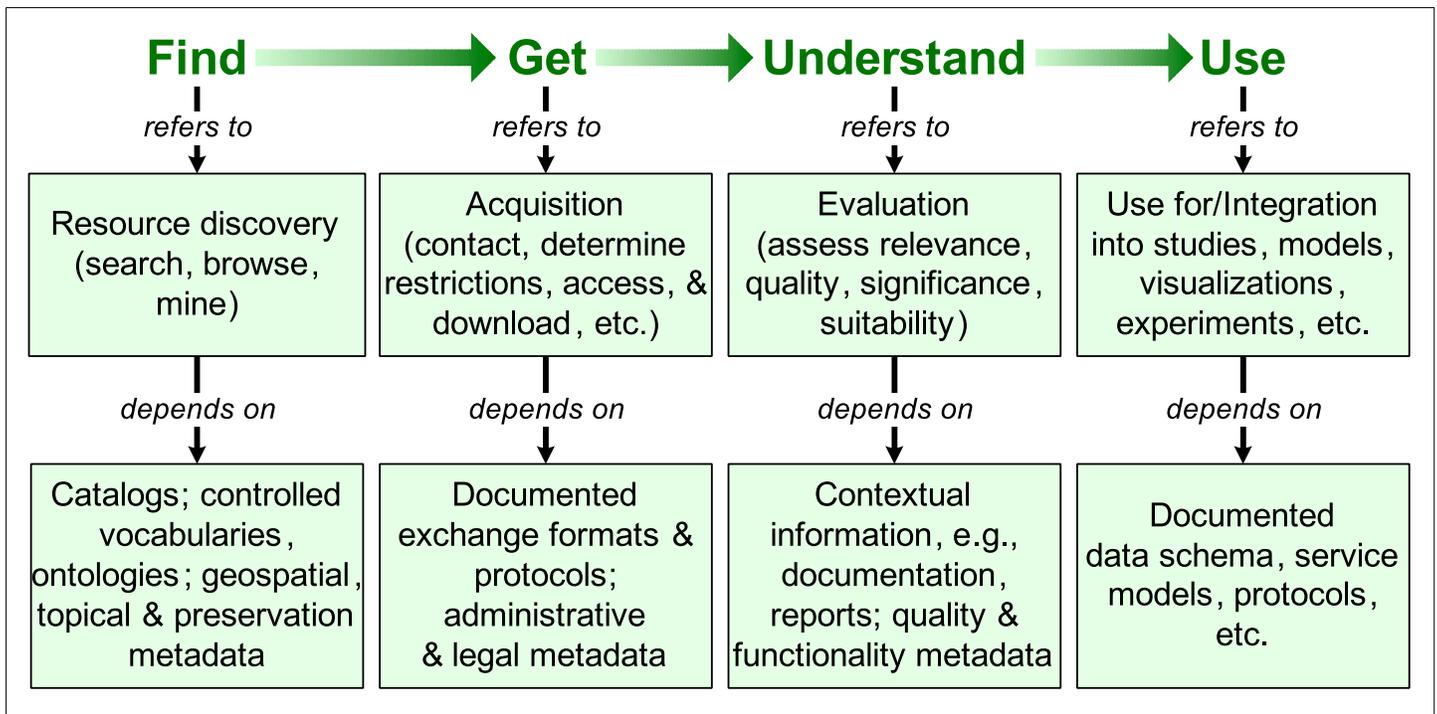
From Tom Gunther and Dave Govoni, USGS, 2006.

# A Vocabulary for Scientific Information and Knowledge Management

## Producer perspective



## Consumer perspective

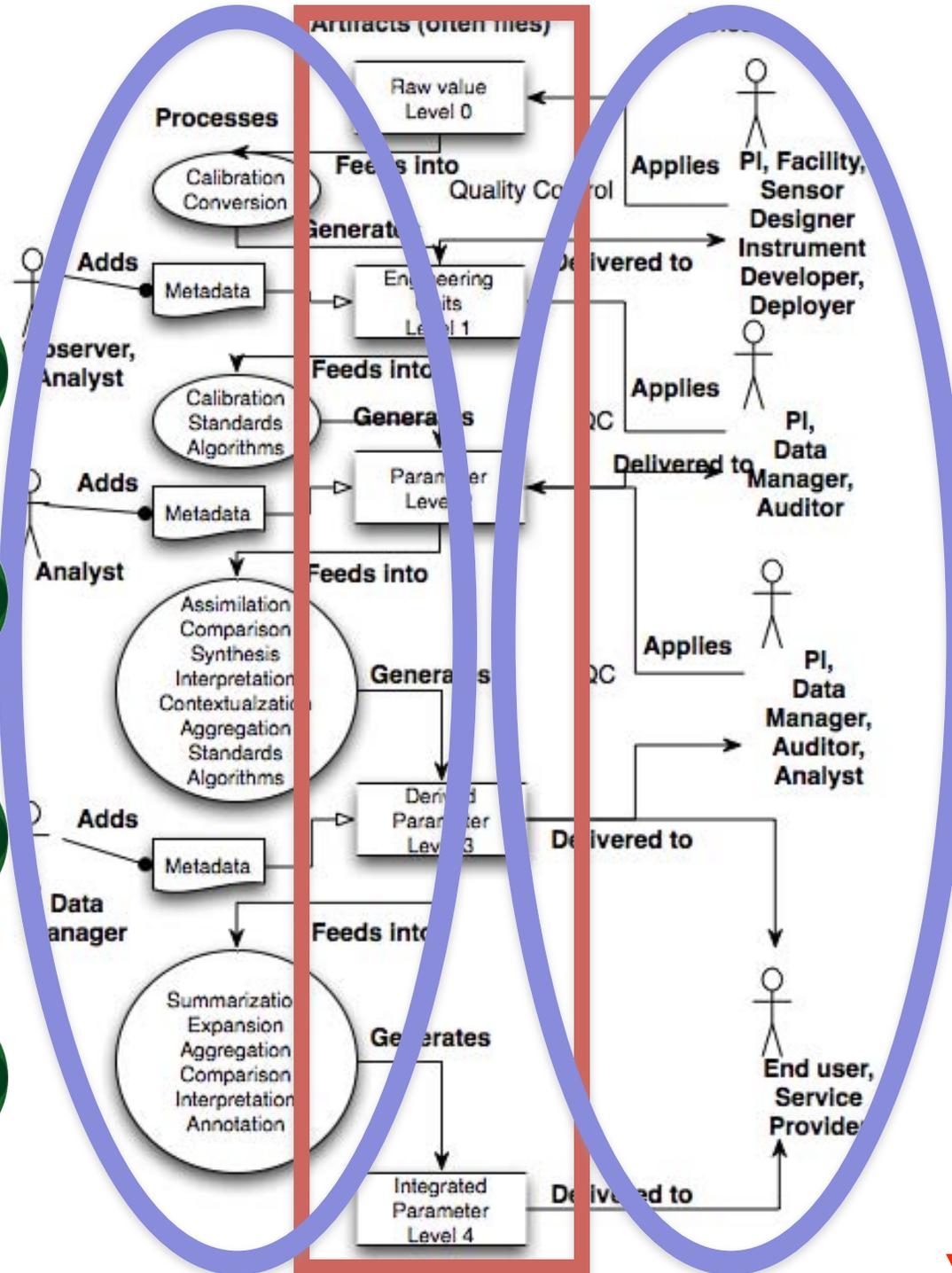
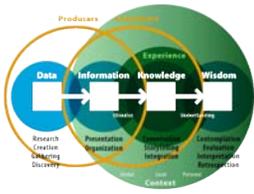
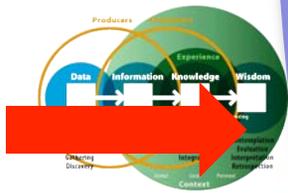
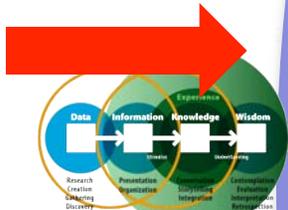
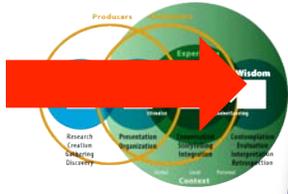


Source: Govoni, D.L. and T.M. Gunther, 2006. Scientific Information Management at the U.S. Geological Survey: Issues, Challenges, and a Collaborative Approach to Identifying and Applying Solutions (Abstract). Geoinformatics 2006 —Abstracts. Scientific Investigations Report 2006-5201, p. 19-20. U.S. Geological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

## **PETER FOX DATA LIFECYCLE DIAGRAMS**

Presented August 2010 to USGS Council for Data Integration meeting.

# Curation stages



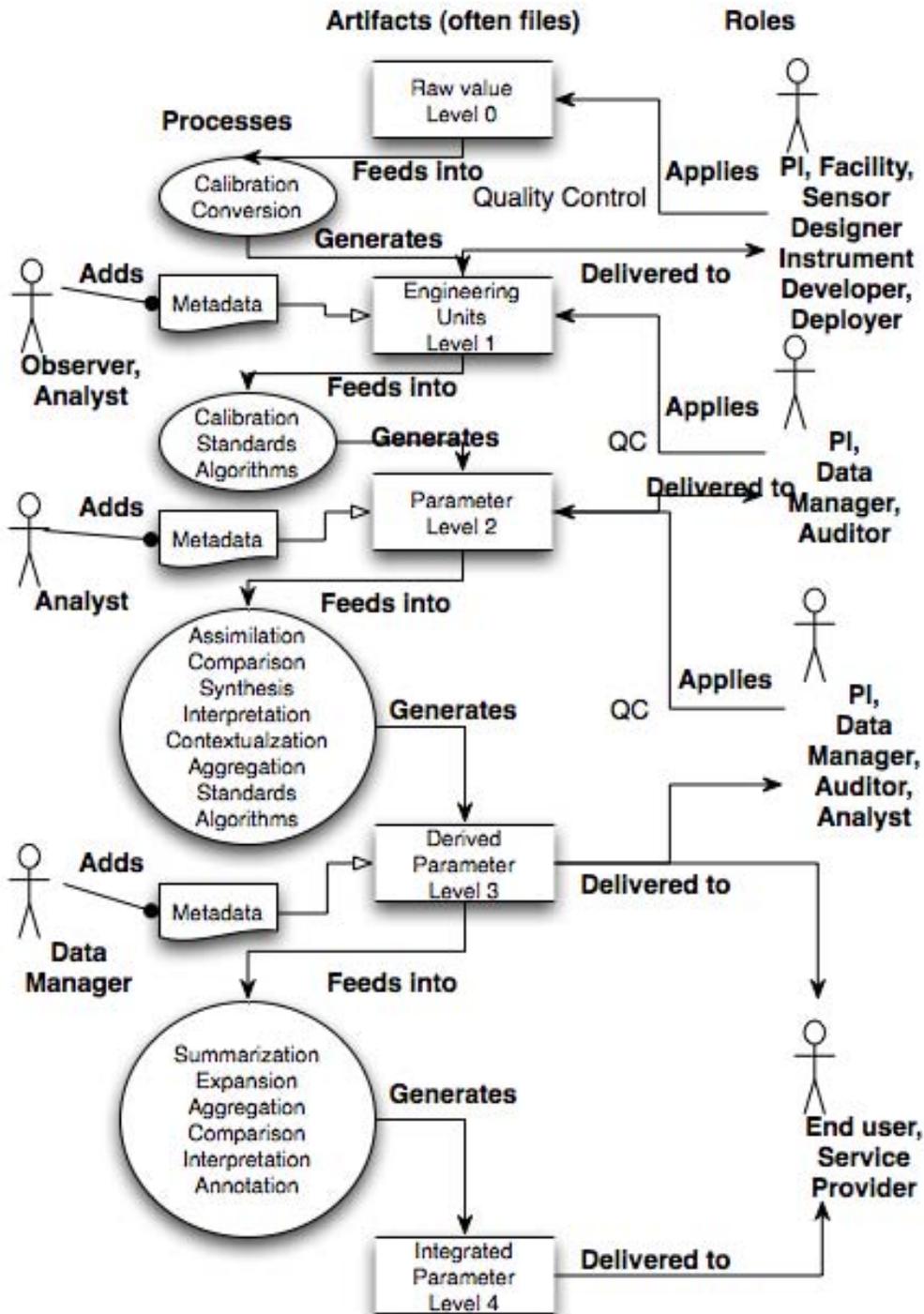
Need to be here

Often 'get' the (meta) data here

Too many 'get' it here



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



## **NATIONAL SCIENCE FOUNDATION DATA LIFECYCLE**

Provided by Vivian Hutchinson via email on 16 November 2010.

This is the National Science Foundation's data lifecycle: (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.