UF | George A. Smathers Libraries

What is a Data Management Plan?

Data Management and Curation Working Group (DMCWG)

UF Institute on Aging, CTRB 3256, 11:30 am – 1:00 pm

Friday, February 15, 2019

UF | George A. Smathers Libraries

Table of Contents

Setting the Context

- "Likewise, data sharing and data management form the foundation of global academic collaboration, discovery and scientific advancement. Sadly, **surveys show that academics rarely get formal training in good data management (let alone best practice), and data management is rarely incentivized by institutions. All too often even the basics are ignored**, with data ending up languishing on a USB stick or on a paper notepad. **If we want research to be discovered, shared, and reused, then the same must be said of the underlying data**." – Scholarly Kitchen (1/30/2019)

Why is data management planning important? (Steneck, 2007, p. 87)

Data management practices should be addressed before any data is collected. Four important issues to take into consideration are:

- ✓ Data Ownership
- ✓ Data Collection
- ✓ Data Storage
- ✓ Data Sharing

What is a data management plan?

A data management plan is designed to encapsulate & articulate details about data from **collection** to **curation** to **preservation** to **dissemination** to **destruction** (data lifecycle).

A data management plan should be an **on-going process** rather than a level of grant requirement for a funding agency program solicitation.

Data Management Models (CCSDS, 2002/2012)?

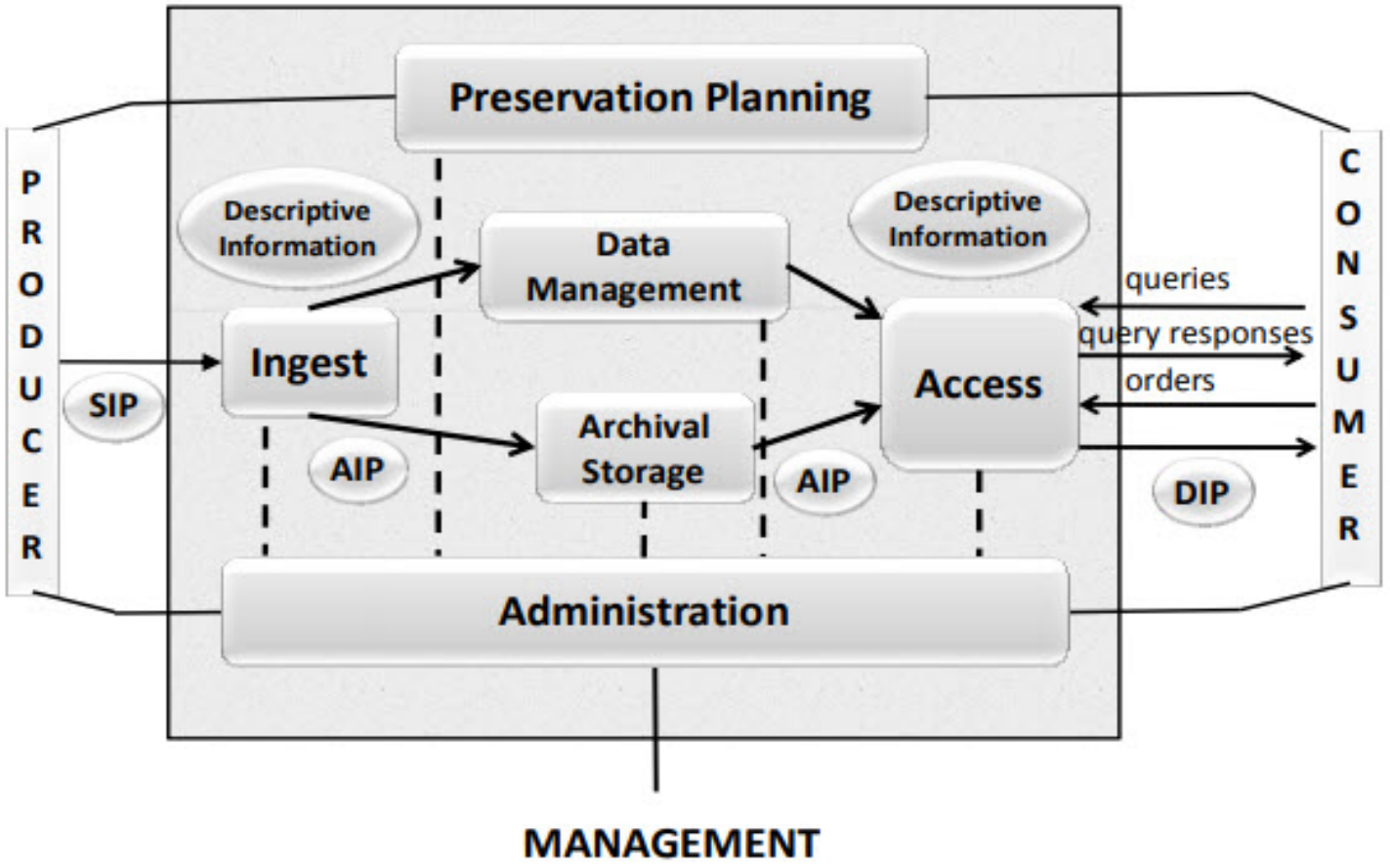

Fig. 1 Open Archival Information System (OAIS) Reference Model (CCSDS, 2002/2012)

Data Management Models (Lord, 2003)?

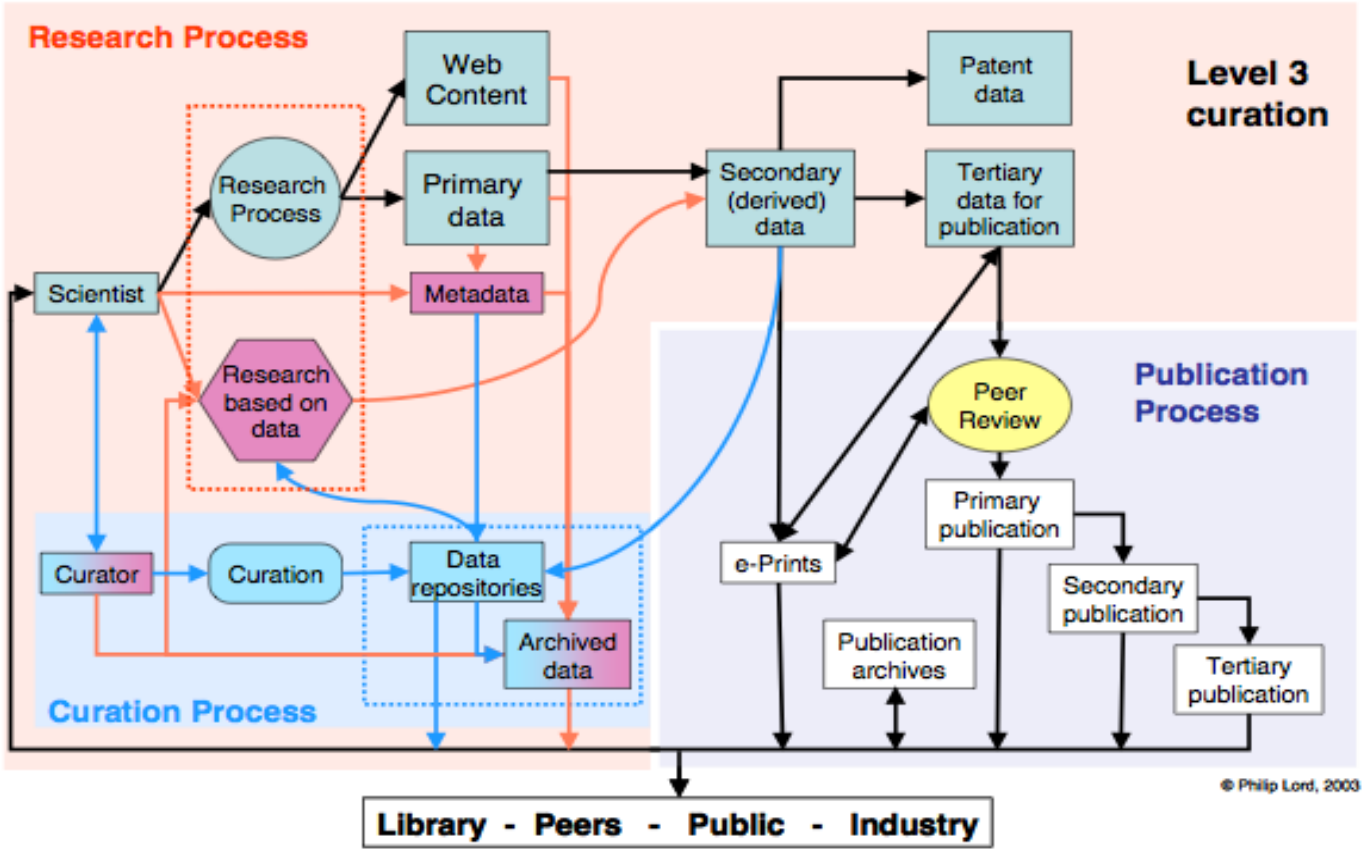

Fig. 2 Level Three Information Flow with Data Curation (Lord, 2003)

Data Management Models (DCC, 2007)?

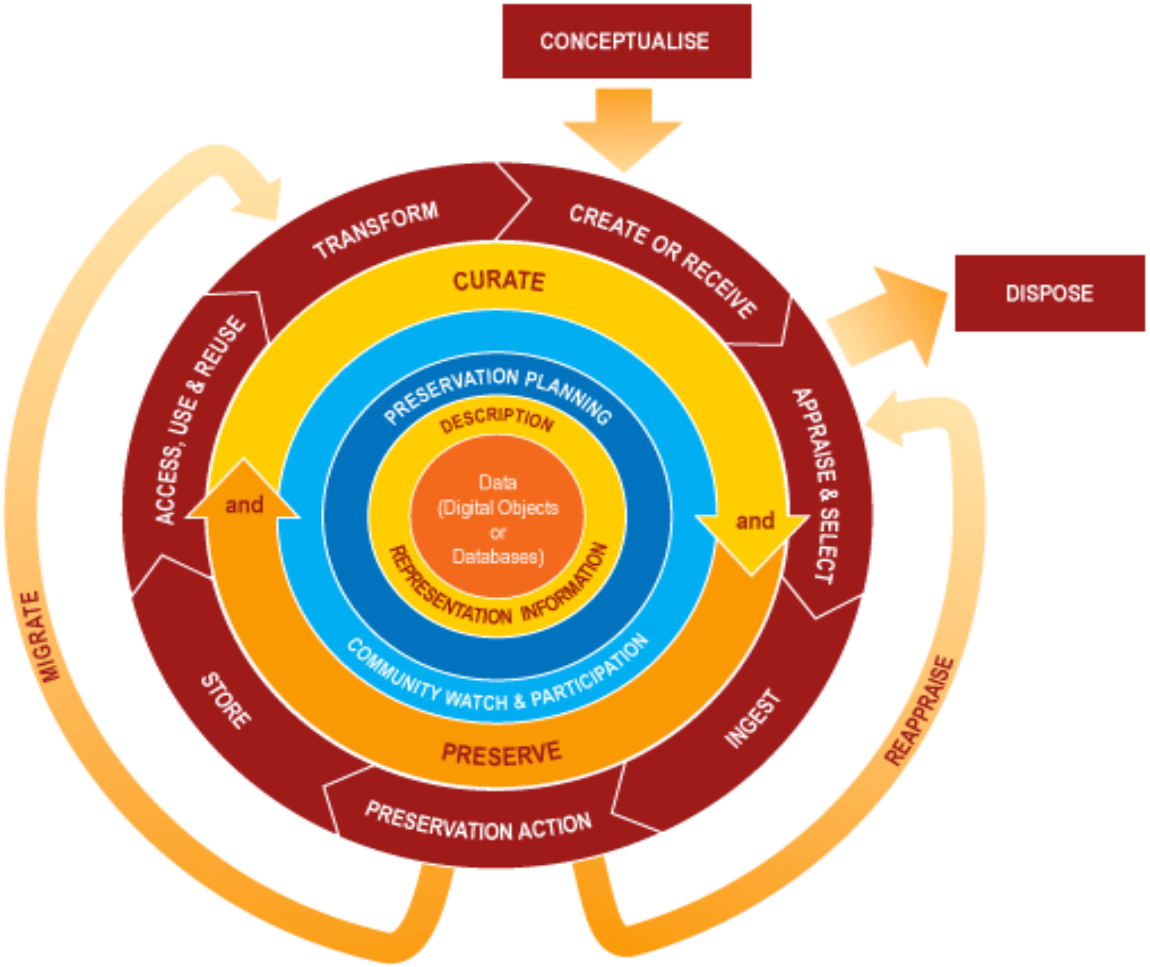

Fig. 3 DCC Curation Lifecycle Model (DCC, 2007)

What is a data management plan?

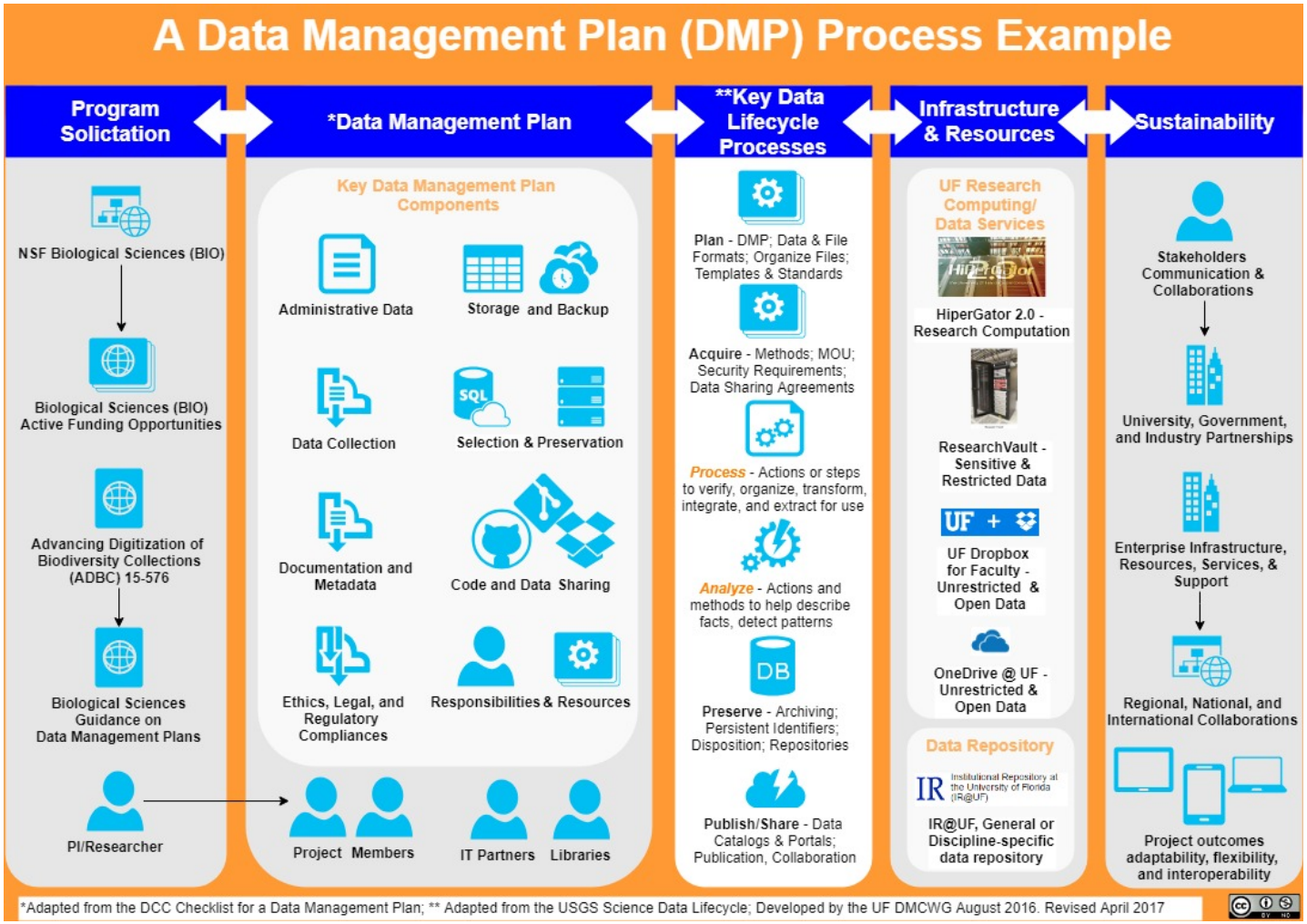

Fig. 4 Data Management Plan Components and Goals

How are Stakeholders involved in data management planning (JISC et al., 2009)?

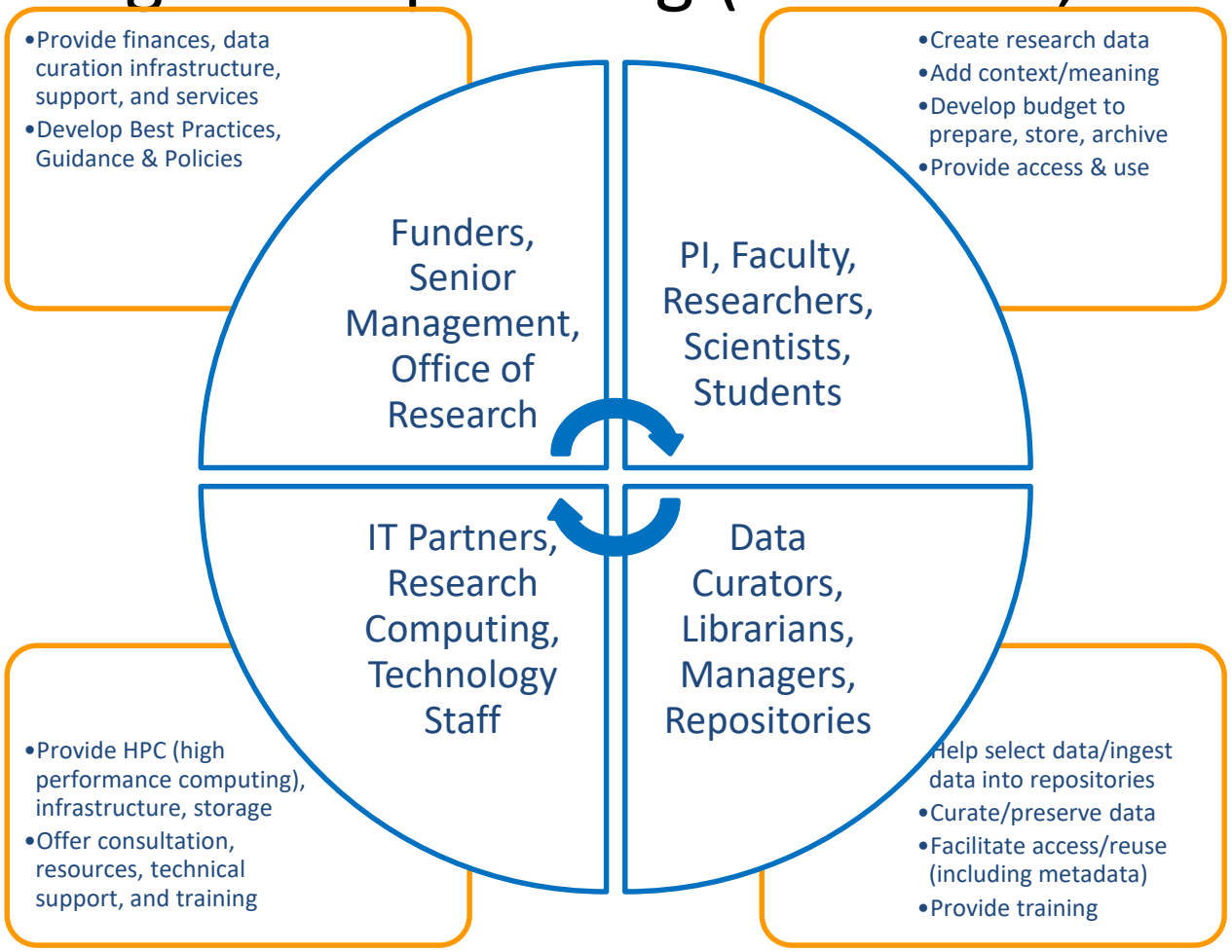

Fig. 5 Stakeholders and Data Management Responsibilities

How are Stakeholders involved in data management planning (cont.)?

Example—Units in Research Data Lifecycle

- Principal Investigator, PI's Research Group, Grants and Contracts Office, Local IT Unit, University Libraries Research Data Services, Sponsor or External Data Provider, Office of Sponsored Programs, Export Compliance, Institutional Review Board, Research Data Security and Compliance Office, High Performance Computing Facility, Purchasing, Risk Management, Office of General Counsel, Privacy Office, Office of Information Security, IT Business Relationship Office, Research Network, IT Cloud Program, Virtual Server Hosting Service, Data Center Services, Identity and Access Management, External Computing Resources, External Data Repositories, Institutional Data Repository, Institutional Scholarship Repository

PennState

Fig. 6 Units involved in data management (used with permission)

UF George A. Smathers Libraries

What are some key components of a data management plan (DCC, 2013)?

Administrative Data

- ID (funder or institution)
- Funder
- Grant Reference #
- Project Name
- Project Description
- PI/Researcher
- Researcher ID (e.g. ORCID)
- Date of 1st version, last update, and related policies

Data Collection

- What data will you collect of create?
  - What type, format, and volume of data? (e.g. text, vcf, 30-50 Gigabyte per dataset)
- How will the data be collected or created?
  - What standards or methodologies will you use?
  - How will you structure and name your folders and files?

UF | George A. Smathers Libraries

What are some key components of a data management plan (DCC, 2013)?

Documentation and Metadata

- What documentation and metadata will accompany the data?
  - What information is needed for the data to be read and interpreted in the future?
  - How will you capture/create the documentation and metadata?
  - What metadata standards will you use and why?

Ethical, Legal, and Regulatory Compliances

- How will you manage any ethical issues?
  - Have you gained consent for data preservation and sharing?
- How will you manage copyright and Intellectual Property Rights (IPR) issues?
  - Who owns the data?
  - How will the data be licensed for reuse?

UF | George A. Smathers Libraries

What are some key components of a data management plan (DCC, 2013)?

Storage and Backup

- How will the data be stored and backed up during research (e.g. Tivoli)?
  - Do you have sufficient storage or will you need to include charges for additional services?
- How will you manage access and security?
  - What are the risks to data security and how will these be managed?

Selection & Preservation

- Which data should be retained, shared, and/or preserved?
  - What data must be retained/destroyed for contractual, legal, or regulatory purposes?
- What is the long-term preservation plan for the dataset?
  - Where e.g. in which repository or archive will the data be held (e.g. NCBI, NCEI)?

UF | George A. Smathers Libraries

What are some key components of a data management plan (DCC, 2013)?

Data Sharing

- How will you share the data?
  - How will potential users find out about your data?
- Are any restriction on data sharing required?
  - What action will you take to overcome or minimize restriction?

Responsibilities & Resources

- Who will be responsible for data management?
  - Who is responsible for implementing the DMP, and ensuring it is reviewed and revised?
- What resources will you require to deliver your plan?
  - Is additional specialist expertise (or training for existing staff) required?

What are some key components of a data management plan (cont.)?

- Examples of DMPs (UF researchers)
  - UF/IFAS http://ufdc.ufl.edu/AA00014835/00088
  - UF/IFAS NCBS http://ufdc.ufl.edu/AA00014835/00111
- UF Research Computing/Data Services
  - HiperGator – High Performance Computing (HPC)
    - 50,000 cores; 3 Petabyte storage
    - Long-term storage per terabyte (orange/blue)
      - https://www.rc.ufl.edu/services/rates/hardware/
  - ResVault
    - Secure data storage and analysis for restricted data
      - HIPAA, ITAR/EAR, Intellectual Property
      - https://www.rc.ufl.edu/services/restricted-data/
  - Dropbox for Faculty, OneDrive @ UF
    - UFIT supported data storage, synchronization and sharing

UF George A. Smathers Libraries

What are some key components of a data management plan (cont.) (Lorenzen et al.*, 2016 – awarded)?

Objective	Output name	Output description	Output (type, format)
Obj. 1	Synthesized data sets	Habitat; Fisheries independent; Fisheries dependent	Habitat (derived, geospatial), Fisheries (derived, tabular)
Obj. 2	Hierarchical analyses of spatial recruitment and angler effort	Reports; Instructions for analyses; Data analyses code; Geospatial images	Reports and Instructions (text, PDF/XML); Code (text, .txt); Geospatial (TIFF and GIS)
Obj. 3	Social-ecological regional system model analyses	Reports; Instructions for analyses; Data analyses code	Reports and Instructions (text, PDF/XML); Code (text, .txt)
Obj. 4	Restoration management strategy evaluation (MSE)	Simulation results; Reports; Instructions for analyses; Data analyses code	Simulation (simulated data, CSV); Reports and Instructions (text, PDF/XML); Code (text, .txt)

Table. 1 Description of project data output and products to be preserved (DMP)*

What are some key data lifecycle processes (USGS, 2013)?

Plan for the data

- Full-lifecycle data management articulation
- Steps to identify and secure resources and utilize infrastructure for data acquisition

Acquire the data

- Collect new data
- Convert/transform legacy data
- Share /exchange data
- Purchase data

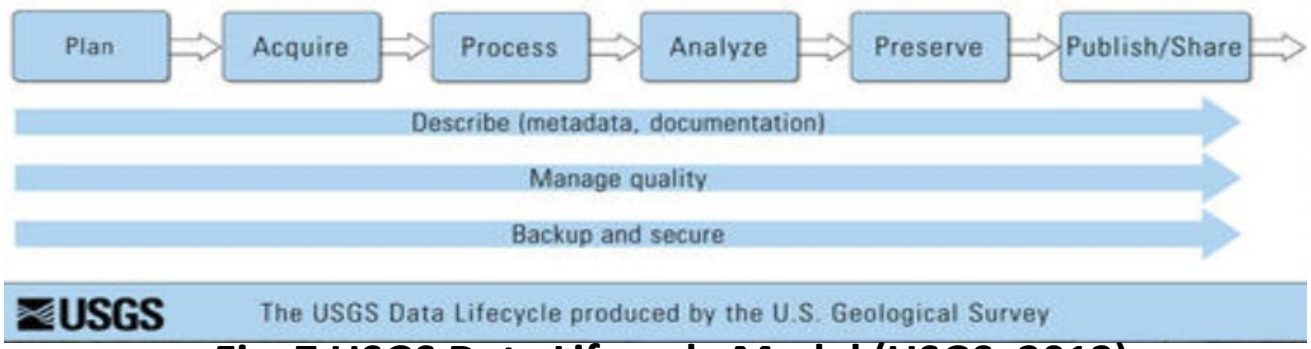

Fig. 7 USGS Data Lifecycle Model (USGS, 2013)

What are some key data lifecycle processes (USGS, 2013)?

Process the data

- Verify, organize, transform, and extract data in an appropriate output for subsequent use

Analyze the data

- Perform actions and method that describe facts, detect patterns, develop explanations, and test hypothesis

What are some key data lifecycle processes (UNSW, 2017)?

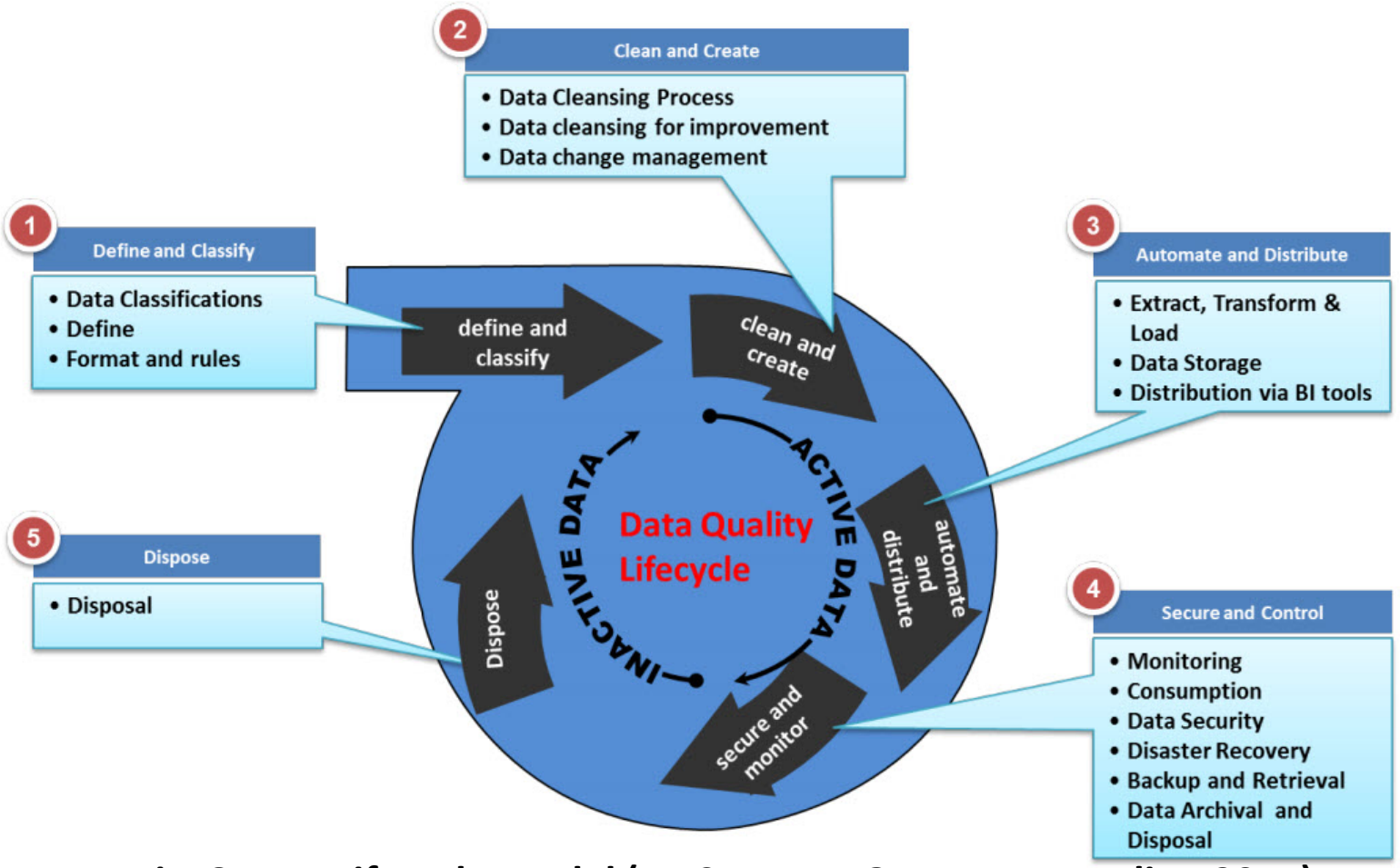

Fig. 8 Data Lifecycle Model (UNSW Data Governance Policy, 2017)

What are some key data lifecycle processes (USGS, 2013)?

Preserve the data

- Perform actions and procedures to keep data for specific period of time for future use (e.g. data retention strategy)

Publish/Share the data

- Process to prepare data for dissemination, public access, and reuse (includes documentation and metadata to facilitate aggregation, dissemination, and representation)

Developing Use Case (UF/IFAS WEC)

- Data Transfer Agreement (collaborative project)
  - PI, Legal Counsel, Data Librarian, collaborators (USF, USGS)
- Data Request
  - Documented collection procedures
  - Documented analytical procedures
  - Quality Assurance (QA) protocol documentation
  - Spatial data
    - Spatial data (i.e. sampling station/site locations and/or project footprint) OR
    - Spatial data (i.e., sampling station/site locations with metadata specifying what parameters are collected where and at what frequency)

UF | George A. Smathers Libraries

References

- CCSDS. (2012). The Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS), Recommended Practice, CCSDS 650.0-M-2, Magenta Book, June 2012. Accessed February 15, 2019 from https://public.ccsds.org/Pubs/650x0m2.pdf. International Organization for Standardization (ISO) Standard ISO 14721:2012.
- DCC. (2013). Checklist for a Data Management Plan. V.4.0. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/data-management-plans.
- DMCWG. (2019). Data Management and Curation Working Group. Accessed online February 15, 2019 from http://ufdc.ufl.edu/AA00014835/00088/allvolumes.
- DMPTool. (2019). https://dmptool.org/.
- ESIP. (2016). Earth Science Information Partners. Data Management Short Course for Scientists. Accessed online Feburary 15, 2019 from http://commons.esipfed.org/datamanagementshortcourse.
- JISC, University of Glasgow – HATII, & DCC. (2009). Data Asset Framework: Implementation Guide. Accessed online February 15, 2019 from http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf.
- Lord, P. and Macdonald, A. (2003). e-Science Curation report, Data curation for e-Science in the UK: an audit to establish requirements for future curation… Accessed online February 15, 2019 from http://tinyurl.com/y7atlrel.
- Lorenzen, K., Camp, E., & Dutka-Gianelli, J. (2016). Synthesizing spatial dynamics of recreational fish and fisheries to inform restoration strategies: red drum in the Gulf of Mexico. Data Management Plan. Accessed online February 15, 2019 from http://ufdc.ufl.edu/AA00014835/00088.
- NSF. (2011). Dissemination and Sharing of Research Results. Accessed online February 15, 2019 from http://www.nsf.gov/bfa/dias/policy/dmp.jsp.
- Scholarly Kitchen. (2019). Guest Post: Encouraging Data Sharing: A small Investment for Large Potential Gain. Accessed online February 15, 2019 from https://tinyurl.com/y4qejzhv.
- Steneck, N. H. (2007). Department of Health and Human Services, Office of Research Integrity (ORI) Introduction to the Responsible Conduct of Research. Accessed February 15, 2019 from http://ori.hhs.gov/sites/default/files/rcrintro.pdf.
- UNSW. (2017). Data Governance Policy. Accessed online February 15, 2019 from https://www.gs.unsw.edu.au/policy/documents/datagovernancepolicy.pdf.
- USGS. (2013). USGS Data Management. Data Lifecycle Overview. Accessed online February 15, 2019 from https://www.usgs.gov/products/data-and-tools/data-management/data-lifecycle.
- USGS CDI. (2012). Community for Data Integration. Data Management Training Accessed online February 15, 2019 from https://my.usgs.gov/confluence/display/cdi/Data+Management+Training.

UF | George A. Smathers Libraries

Thank you

Questions/comments

Contacts:

Plato Smith, Data Management Librarian, plato.smith@ufl.edu

Data Management and Curation Working Group

DATAMGMT-L@LISTS.UFL.EDU