Developing Data Management Education, Support, and Training

Plato Smith, Data Management Librarian

University of Florida

Librarian's Role in the Research

March 20, 2018

Setting the Context

"A record if it is to be useful to science, must be continuously extended, it must be stored, and above all it must be consulted." – Dr. Vannevar Bush (July 1945, As We May Think)

"The process by which data is captured and maintained continues to evolve and mature as scientific needs change." – DAF Interview P1 Participant (2013, Q18)

Table of Contents

1. Linking stakeholders, liaisons, and students
2. Articulating a data management plan
3. Using the OAIS Model to Explain Concepts
4. Explaining some key components of a data management plan
5. What are some key data lifecycle processes?
6. How can you develop a data management plan?
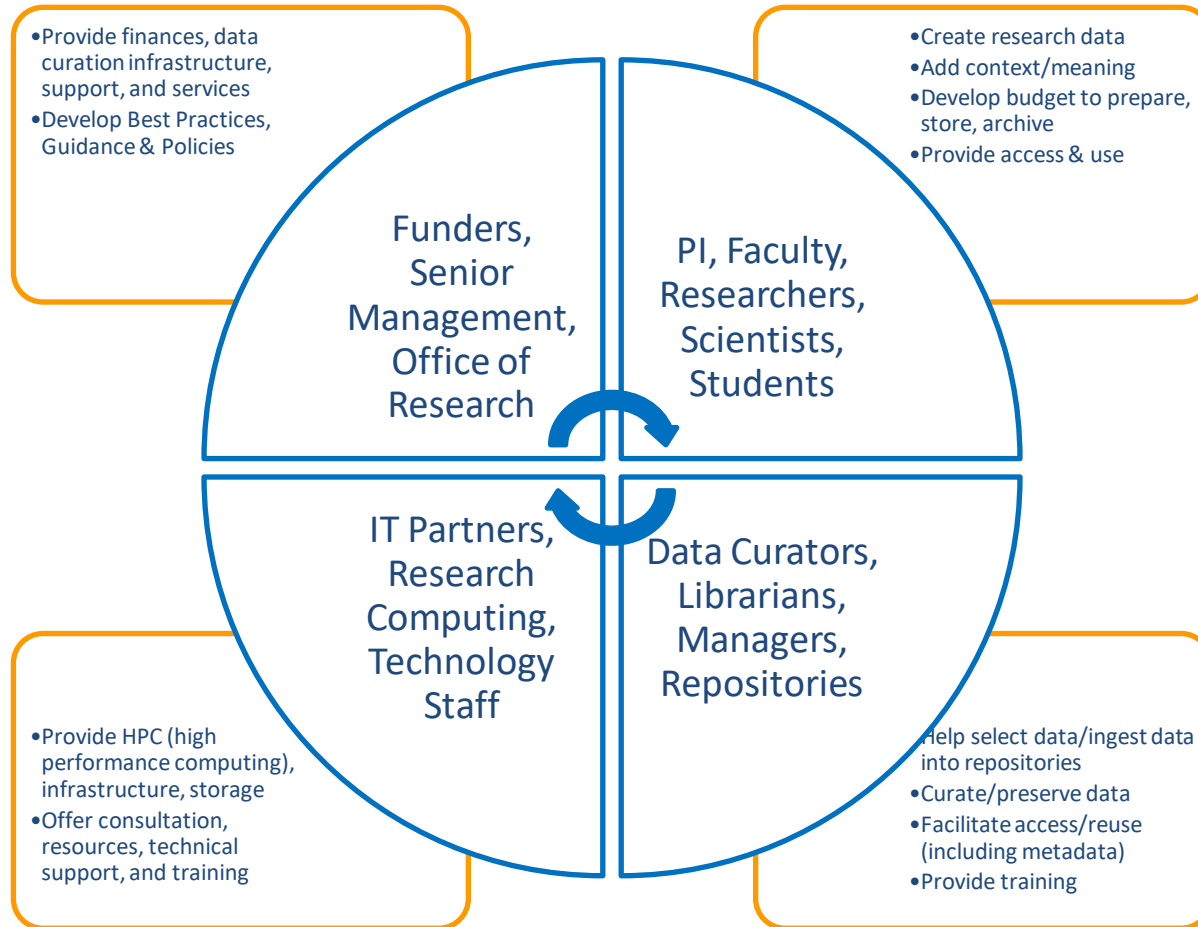7. Resources

Linking stakeholders, liaisons, and students

- Provide finances, data curation infrastructure, support, and services
- Develop Best Practices, Guidance & Policies

- Create research data
- Add context/meaning
- Develop budget to prepare, store, archive
- Provide access & use

Funders, Senior Management, Office of Research

PI, Faculty, Researchers, Scientists, Students

IT Partners, Research Computing, Technology Staff

Data Curators, Librarians, Managers, Repositories

- Provide HPC (high performance computing), infrastructure, storage
- Offer consultation, resources, technical support, and training

- Help select data/ingest data into repositories
- Curate/preserve data
- Facilitate access/reuse (including metadata)
- Provide training

Fig. 1 Stakeholders and Data Management Responsibilities

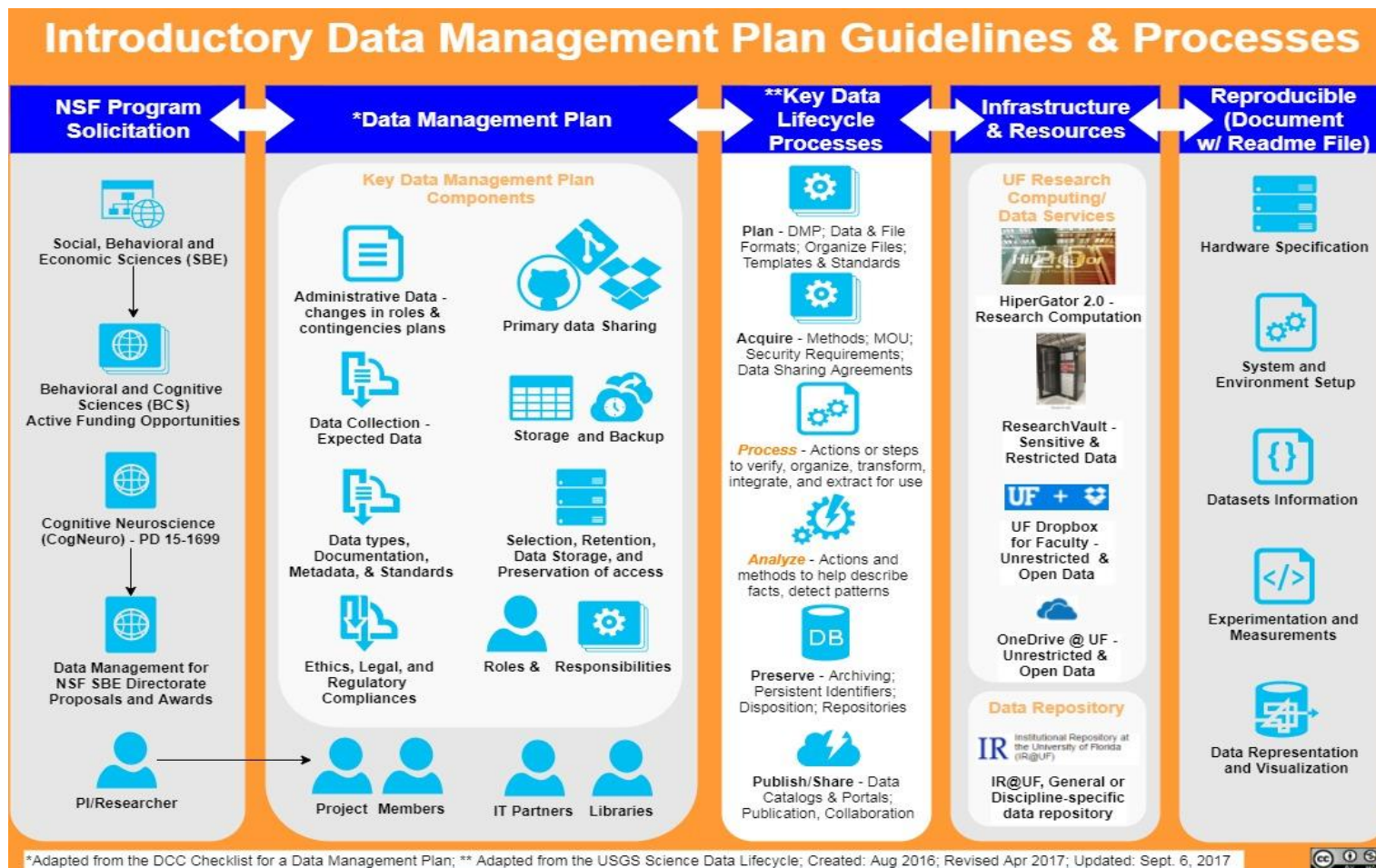# Articulating a data management plan

Fig. 2 Data Management Plan Components and Goals
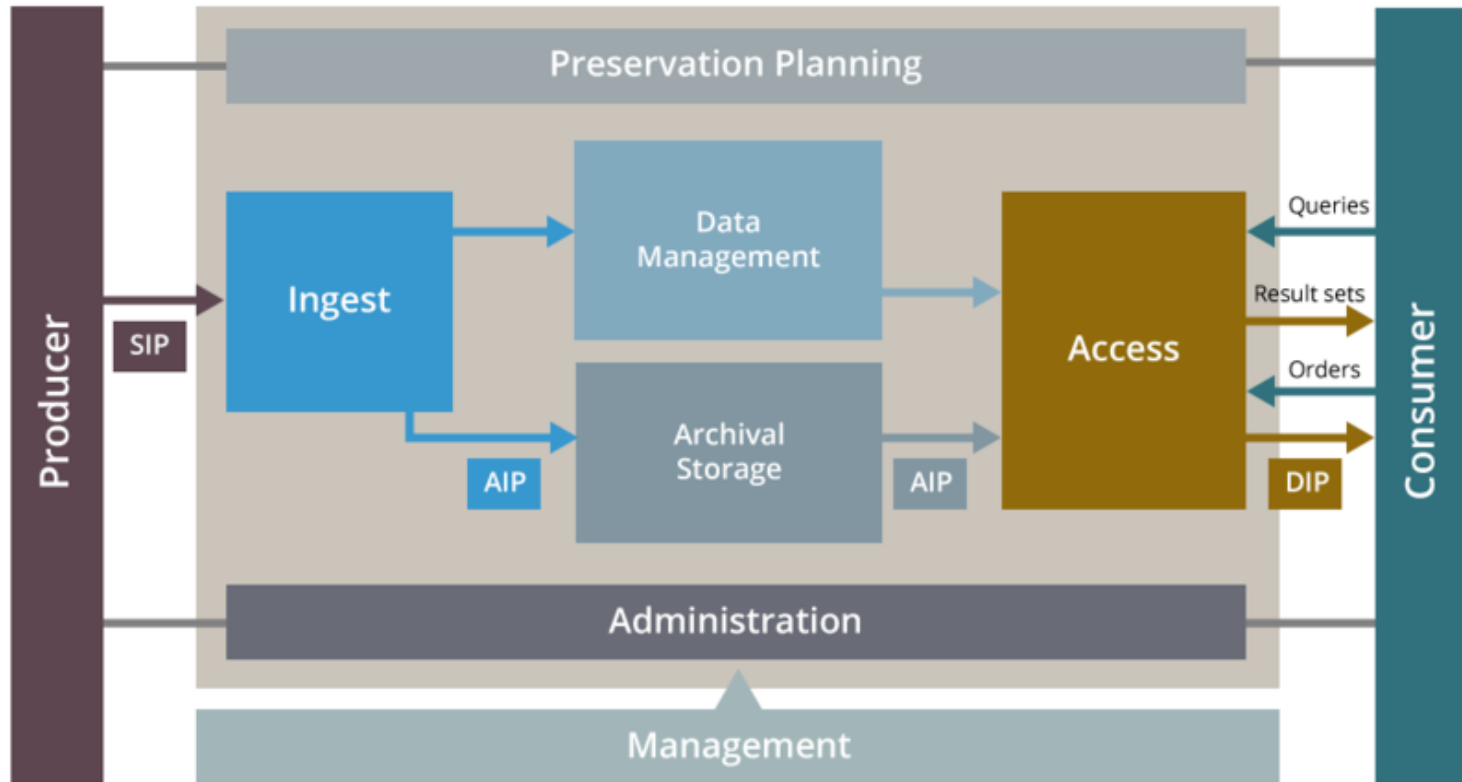
Using the OAIS Model to Explain Concepts

Fig. 3 Open Archival Information System (OAIS) Model (cessda, 2018)

Explaining 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?

Explaining 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?

Explaining 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. FDA, 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)?

Explaining 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?
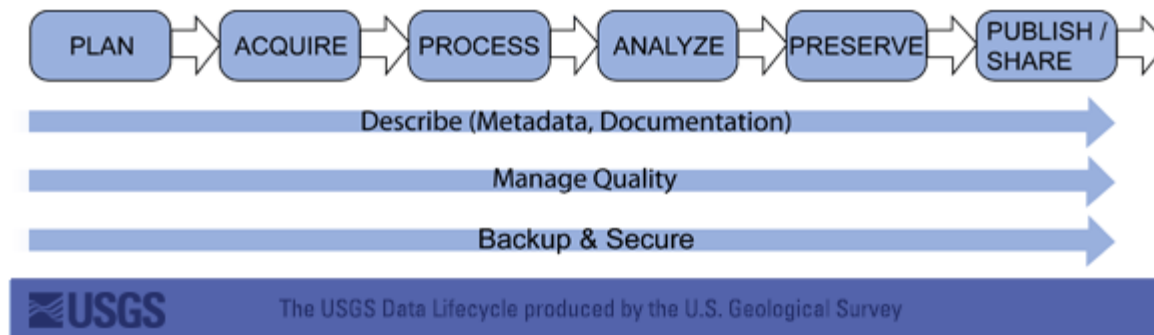 - Is additional specialist expertise (or training for existing staff) required?

What are some key 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

PLAN → ACQUIRE → PROCESS → ANALYZE → PRESERVE → PUBLISH / SHARE →

Describe (Metadata, Documentation)

Manage Quality

Backup & Secure

USGS The USGS Data Lifecycle produced by the U.S. Geological Survey

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 (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)

How can you develop a data management plan?

- Hand on exercise portion of this training:
  - Navigate to https://dmptool.org
  - Click on Login in upper-right hand corner
  - Click on dropdown arrow to select your institution
  - Select the University of Florida
  - Click on the Next button
  - Login with you GatorLink credentials
  - Click on Create New DMP
  - Select DMP Template from DCC

References

- cessda. (2018). Tutorial: OAIS. Retrieved March 20, 2018 from https://www.cessda.eu/Projects/All-projects/CESSDA-SaW/WP4/Tutorial-OAIS.
- 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
- DMPTool. (2018). Retrieved March 20, 2018 from https://dmptool.org/
- ISO. (2013). ISO/IEC 27001 – Information security management. Retrieved March 20, 2018 from http://www.iso.org/iso/iso27001.
- JISC, University of Glasgow – HATII, & DCC. (2009). Data Asset Framework: Implementation Guide. Retrieved March 20, 2018 from http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf.
- NSF. (2011). Dissemination and Sharing of Research Results. Retrieved March 20, 2018 from http://www.nsf.gov/bfa/dias/policy/dmp.jsp.
- USGS. (2013). USGS Data Management. Data Lifecycle Overview. Retrieved March 20, 2018 from https://www2.usgs.gov/datamanagement/why-dm/lifecycleoverview.php.