Data Management Plan (DMP) Training
# Table of Contents

1. Why is data management planning important?  
2. Who are the Stakeholders involved in data management planning?  
3. What is a data management plan?  
4. What are 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. References  
8. Contacts  
9. Acknowledgements
Why is data management planning important?

2011 NSF Data Management Plan Requirements

2013 Office of Science and Technology Policy Policy

2015 NSF Public Access Plan (NSF 15-52)
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:

1. Data Ownership
2. Data Collection
3. Data Storage
4. Data Sharing
   - Data management plan for a facility
   - De-identifying sensitive data
   - Permanent data deposit
   - Reproducibility of data
Who are the Stakeholders involved in data management planning? (JISC et al., 2009, p. 3)

Fig. 1 Stakeholders and Data Management Responsibilities
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.
What is a data management plan? (Jones, et al., 2013, p. 5)

Fig. 2 Components of RDM support services (used with permission)
What is a data management plan? (ANDS, 2016)

Elements of data management plans

The diagram below sets out the various elements researchers and research institutions may wish to consider when developing data management plans. Each element is linked to further information.

Fig. 3 Elements of data management plans (used with permission)
What is a data management plan?

**A Data Management Plan (DMP) Process Example**

**Program Solicitation**
- NSF Biological Sciences (BIO)
- Biological Sciences (BIO) Active Funding Opportunities
- Advancing Digitization of Biodiversity Collections (ADBC) 15-576
- Biological Sciences Guidance on Data Management Plans
- PI/Researcher

**Data Management Plan**
- Key Data Management Plan Components
  - Administrative Data
  - Storage and Backup
  - Data Collection
  - Selection & Preservation
  - Documentation and Metadata
  - Code and Data Sharing
  - Ethics, Legal, and Regulatory Compliances
  - Responsibilities & Resources
- Project Members
- IT Partners
- Libraries

**Key Data Lifecycle Processes**
- Plan - DMP, Data & File Formats, Organize Files, Templates & Standards
- Acquire - Methods; MOC; Security Requirements, Data Sharing Agreements
- Process - Actions or steps to verify, organize, transform, integrate, and extract for use
- Analyze - Actions and methods to help describe facts, detect patterns
- Preserve - Archiving; Persistent Identifiers; Disposition; Repositories
- Publish/Share: Data Catalogs & Portals; Publication, Collaboration

**Infrastructure & Resources**
- UF Research Computing/ Data Services
  - HiperGator 2.0
  - ResearchVault
  - GatorBox
  - UF Dropbox for Faculty
  - OneDrive @ UF
- Data Repository
  - Intellectual Repository at the University of Florida
  - IR@UF: General or Discipline specific data repository

**Sustainability**
- Stakeholders Communication & Collaborations
- University, Government, and Industry Partnerships
- Enterprise Infrastructure, Resources, Services, & Support
- Regional, National, and International Collaborations
- Project outcomes: adaptability, flexibility, and interoperability

*Adapted from the DCC Checklist for a Data Management Plan; **Adapted from the USGS Science Data Lifecycle. Developed by the UF DMCWG August 2016.*

Fig. 4 Data Management Plan Components and Goals
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](https://orcid.org))
- ✓ Date of 1\textsuperscript{st} version, last update, and related policies

**Data Collection**
- ❑ What data will you collect or 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?
What are some key components of a data management plan?

<table>
<thead>
<tr>
<th>Data Classification</th>
<th>Data type</th>
<th>Justification</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted (sensitive)</td>
<td>Student records (non-directory)</td>
<td>FERPA</td>
<td>Data that if breached owning to accidental or malicious activity would have a high impact on the University’s activities and objectives.</td>
</tr>
<tr>
<td>Restricted (highly sensitive)</td>
<td>Credit card cardholder data</td>
<td>PCI, FS 317.5881, FS 501.171</td>
<td></td>
</tr>
<tr>
<td>Restricted (highly sensitive)</td>
<td>Patient medical records (identifiable)</td>
<td>HIPAA</td>
<td>This classification describes the intended audience from a restricted UF organizational unit or perspective. Dissemination is based on strict academic, research or business need.</td>
</tr>
<tr>
<td>Restricted</td>
<td>Patient billing records</td>
<td>HIPAA</td>
<td></td>
</tr>
<tr>
<td>Restricted</td>
<td>Social Security Numbers²</td>
<td>FS 317.5881</td>
<td></td>
</tr>
<tr>
<td>Restricted</td>
<td>Export Controlled data</td>
<td>ITAR, EAR</td>
<td></td>
</tr>
<tr>
<td>Sensitive (private)</td>
<td>Animal research protocols</td>
<td>Competitive and commercial potential</td>
<td>Data that if breached owning to accidental or malicious activity would have a high impact on the University’s activities and objectives.</td>
</tr>
<tr>
<td>Sensitive</td>
<td>System security plans</td>
<td>Protective information</td>
<td></td>
</tr>
<tr>
<td>Sensitive</td>
<td>Unpublished research results or other intellectual property (IP)</td>
<td>Competitive and commercial potential</td>
<td>This classification describes the intended audience from a restricted UF organizational unit or perspective. Dissemination is based on strict academic, research or business need.</td>
</tr>
<tr>
<td>Sensitive</td>
<td>Exams (material, questions, and results)</td>
<td>Exam integrity</td>
<td></td>
</tr>
<tr>
<td>Sensitive</td>
<td>Employee data (not including SSN)</td>
<td>Employee privacy</td>
<td></td>
</tr>
<tr>
<td>Open (unclassified)</td>
<td>UF Directory (students &amp; staff)</td>
<td>Intended for public use with all data restricted by FERPA removed</td>
<td>Data that if breached owning to accidental or malicious activity would have an insignificant impact on the University’s activities and objectives.</td>
</tr>
<tr>
<td>Open</td>
<td>University regulations</td>
<td>Intended for public use</td>
<td></td>
</tr>
<tr>
<td>Open</td>
<td>Course catalog</td>
<td>Intended for public use</td>
<td></td>
</tr>
<tr>
<td>Open</td>
<td>Public web sites</td>
<td>Intended for public use</td>
<td></td>
</tr>
<tr>
<td>Open</td>
<td>De-identified patient information²</td>
<td>Intended for public use with all data restricted by HIPAA removed</td>
<td>This label describes the intended audience.</td>
</tr>
</tbody>
</table>

Fig. 5 UF Data Classification Guidelines (revised 12/2016)
What are some key components of a data management plan? (example)

**Administrative Data**
- ID: UFID#12345678
- PI: Dr. Smith
- UF Libraries Strategic Opportunity Program
- SO48 – Grant, 7/2016
- Documenting VCF Datasets Use Case

**Data Collection**
- Format
  - RID: N-9123-2015
  - 2016-08-08, v.1
  - Last updated: 2017-01-11
- Software
  - Long-term access
- Sharing
What are some key components of a data management plan? (example)

Fig. 6 Raw data to SRA, results to journal, vcf to zenodo and IR@UF workflows
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?

**Ethics, 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?
What are some key components of a data management plan? (Lorenzen et al., 2016, Table 6)

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

**Data Sharing**
- Secured environment (e.g. restricted or sensitive data)
- Discipline-specific data repository (e.g. re3data)
- General data repository (e.g. OAIS; ISO 14721; Data Seal or Approval)
- UF Institutional Repository (IR@UF)

**Responsibilities & Resources**
- PI (DMP Implementer)
- Research Assistant, Project Manager, Postdoc, IT Support, Graduate Student
- Policies compliancy
- Budget/costs/funding
- Collaborations, partners, consortiums (e.g. idigbio)
- Agreements/MOUs/Policies
What are some key components of a data management plan? (UF resources)

- **UF Research Computing/Data Services**
  - **HiperGator** – High Performance Computing (HPC)
    - 50,000 cores
    - 3 Petabytes storage
  - **ResVault** (FISMA moderate rating)
    - Secure data storage and analysis for restricted data
      - HIPAA, ITAR/EAR, FERPA, Intellectual property data
  - **Gatorbox, Dropbox for Faculty, OneDrive @ UF**
    - UFIT supported data storage, synchronization and sharing
    - [https://www.rc.ufl.edu/services/data-management/](https://www.rc.ufl.edu/services/data-management/)
What are some key components of a data management plan? (tools)

Data Management Plan Tools
- DMPTool, Public DMPs, United States – https://dmptool.org/
- DMP online, Digital Curation Centre, United Kingdom - https://dmponline.dcc.ac.uk/
- DIRT Digital Research Tools - https://dirtdirectory.org/
- NOAA. Archiving your Data at NCEI, https://www.ncdc.noaa.gov/customer-support/archiving-your-data-ncdc
- ORNL DAAC, Distributed Active Archive Center for Biogeochemical Dynamics Tools, https://daac.ornl.gov/tools.shtml
- UCAR, Unidata Software – Data Visualization & Analysis Software; Data Access & Management Software - http://www.unidata.ucar.edu/software/
What are some key data lifecycle processes?  
(USGS, 2013)

Plan for the data

- Full-lifecycle data management articulation
  - Ownership, Collection, Storage, Sharing

- Steps to identify and secure resources and utilize infrastructure for data acquisition
  - Resources, Support, & Tools

Acquire the data

- How will you collect data?
  - Collect new data
  - Convert/transform legacy data
  - Share /exchange data
  - Purchase data
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?

(Lorenzen et al., 2016 - DMP example – successfully awarded – refer to handout)

• **Project Title**: Synthesizing spatial dynamics of recreational fish and fisheries to inform restoration strategies: red drum in the Gulf of Mexico

• **Funding Agency**: A Data Synthesis Project supported by the National Academies Gulf Research Program

• **Funding Support**: Total budget requested: $480, 248

• **Key Contacts:**
  - **Kai Lorenzen** (Project Director, PI), University of Florida (UF), School of Forest Resources and Conservation (SFRC), Fisheries and Aquatic Sciences Program (FAS)
  - **Ed Camp** (PostDoc, Co-PI), UF, SFRC, FAS (modeling and quantitative analysis)
  - **Jynessa Dutka-Gianelli** (PostDoc, Co-PI), UF, SFRC, FAS (data management)
How can you develop a data management plan?

Hands-on exercise portion of this workshop:

- **Navigate** to [https://dmptool.org](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 your GatorLink credentials
- **Click on** Create New DMP
- **Select** DMP Template from DCC
References


Contacts

✓ Data Management and Curation Working Group, DATAMGMT-L@LISTS.UFL.EDU
  ➢ UF Libraries Subject Specialists, Faculty Liaisons
  ➢ Plato Smith, Data Management Librarian, plato.smith@ufl.edu

✓ UF Research Computing Information Technology, https://www.rc.ufl.edu/contact/
Acknowledgements

• Dr. George Michailidis, Founding Director, UF Informatics Institute
• Alethea Geiger, Administrative Specialist III, UF Informatics Institute
• Flora Marynak, Office Assistant, UF Informatics Institute
• Data Management and Curation Working Group
• UF Libraries, UF Division of Sponsored Programs, UF Research Computing, UF/IFAS, UF Department of Molecular Genetics & Microbiology
Thank you, my friends!

Questions and Comments

Plato Smith

plato.smith@ufl.edu