

Research Data Management

*Dalal Rahme
Data Services Librarian
American University of Beirut*

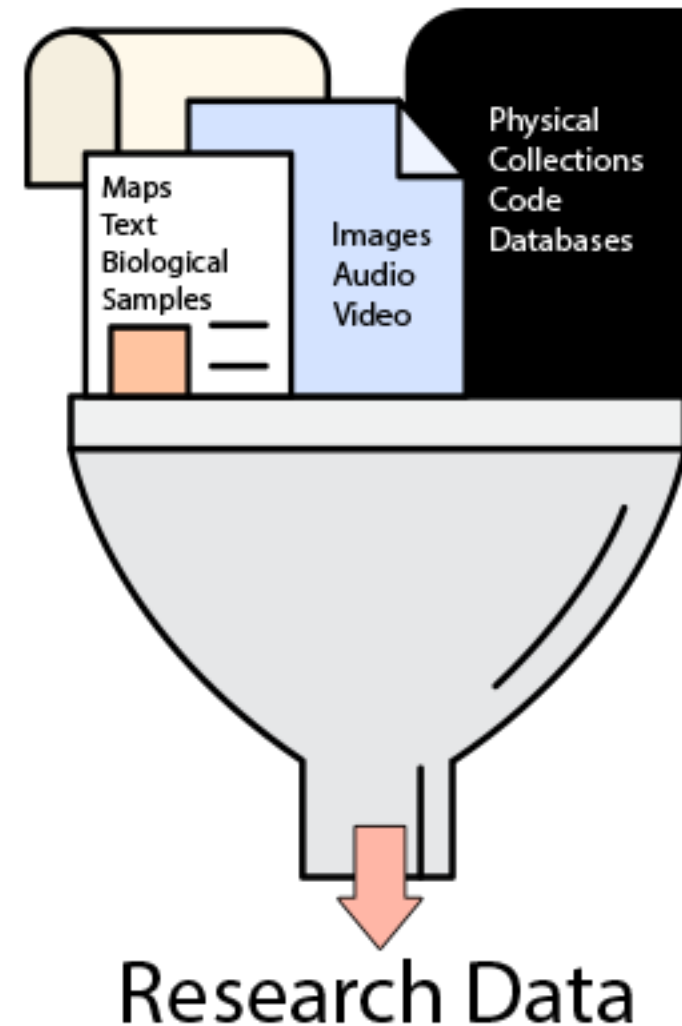
Outline

1. Research Data
2. Video: the problem
3. Components of a Data Management Plan
4. Exercise: Write a DMP



Research Data

“Recorded **factual material** commonly accepted in the scientific community as necessary to **validate** research findings.” (*NIH, 2003*)



Research Data

- All original sources or material (digital or not) created or collated to conduct a research project.
- All data that is analysed to answer the research question(s).
- Research data “may be raw, abstracted or analysed, experimental or observational.” ([UCL Research Data Policy](#))
- Examples include: questionnaires, audio recordings, database entries, blood samples, lab notebooks, list of measurements, photographs, maps etc.

Data Sharing and Management Snafu in 3 Short Acts
by Karen Hanson, Alisa Surkis & Karen Yacobucci
NYU Health Sciences Libraries
August 3, 2012 (Last Update: December 12, 2012)



What seems to be the problem?



The problem

Understanding the issue:

“Everything you need to know is in the article.”

Location and availability of data:

“So many boxes....”

Readability of data:

“I cannot read hexadecimal!”

Documentation:

“My co-author knows what the content of Sam2 is... and Sam3 ... and Sam4.”

Sharing Research Data

- promote innovation and potential new data uses
- maximize transparency and accountability
- encourage the improvement and validation of research methods
- reduce the cost of duplicating data collection
- increase the impact and visibility of research
- promote the research that created the data and its outcomes
- provide a direct credit to the researcher as a research output in its own right



Making Data Fair

Findable Persistent identifier (like a DOI or Handle)

- Rich metadata for description
- Available via discovery portals

Accessible

- Clarity and transparency around the conditions governing access and reuse

Interoperable

- Use of community agreed formats, language and vocabularies for data and metadata

Reusable

- Rich contextual information
- Information on how data was formed
- Machine readable metadata and licensing information

Research Data Lifecycle



Research Data Management

- Involves curation and preservation of both data and associated metadata
- Extends beyond the project life-cycle (long-term) to ensure sustained accessibility and re-use
- Requires effective planning – i.e. Research Data Management Plan (RDMP or DMP)



What is a Data Management Plan(DMP)?



A Data Management Plan

- A short document
- A living document
- Required by many funding agencies
- Includes 2 topics:
 - ✓ What data will your research generate?
 - ✓ What is your plan for managing the data?



Some of the Funders requiring a DMP



Arts & Humanities
Research Council



BBSRC
bioscience for the future

EPSRC
Pioneering research
and skills

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

MRC | Medical
Research
Council

DFID Department for
International
Development

NERC
SCIENCE OF THE
ENVIRONMENT



Science & Technology
Facilities Council

**breast
cancer
campaign**
Research that saves lives



**CANCER
RESEARCH
UK**

welcometrust



**World Health
Organization**



European Research Council
Established by the European Commission



**BILL & MELINDA
GATES foundation**

NIH National Institutes of Health
Turning Discovery Into Health

NHS
**National Institute for
Health Research**

General Components of a DMP

1. Data Products
2. Data Description
3. Access and Security
4. Policies for re-use and sharing
5. Plans for archiving and preservation

I. Data Production

What data will you collect or create?

- Observational
- Experimental
- Simulation
- Derived
- Reference

- Word Documents
- Spreadsheets
- Lab Notebooks
- Field Notebooks
- Diaries
- Audio
- Video
- Photographs
- Films
- Slides
- Artifacts
- Specimens
- Databases
- Logs
- Models
- Scripts/Code
- Workflows
- SOPs
- Metadata

I. Data Production

How will the data be collected or created? **Methodology**



Formats

In which **format** the data will be kept?

Proprietary
or
Non-Proprietary

Proprietary format

- Is a file format that a company owns and controls.
- Data in this format may need proprietary software to be read reliably.
- The description of the format may be confidential or unpublished, and can be changed by the company at any time.
- Proprietary software usually reads and saves data in its own proprietary format.
- For example, different versions of Microsoft Excel use the proprietary XLS and XLSX formats.

Non proprietary file formats

Tables, spreadsheets, and databases	<ul style="list-style-type: none">• Portable Document Format (.pdf)• Tab-separated tables (.txt – sometimes .tsv or .tab)• Comma-separated tables (.csv or .txt)• Other standard delimiter (e.g. colon, pipe)• Fixed-width• OpenDocument Spreadsheet (.ods)• OpenDocument Database (.ods)
Multidimensional arrays	<ul style="list-style-type: none">• Common Data Format (.cdf)• Network Common Data Format (.nc)• Hierarchical Data Format (usually .hdf or .h5)
Statistics	<ul style="list-style-type: none">• Data Documentation Initiative (.xml)
Web data	<ul style="list-style-type: none">• Javascript Object Notation (.json)• Extensible Markup Language (.xml)• Hypertext Markup Language (.html)
Image files	<ul style="list-style-type: none">• JPEG (.jpg or .jp2)• Portable Network Graphics (.png)• Scalable Vector Graphics (.svg)• TIFF (.tiff or .tif)• Portable Document Format (.pdf)

Non proprietary file formats

Sound files	<ul style="list-style-type: none">• WAVE (.wav)• FLAC (.flac)• MPEG-3 (mp3—may not be suitable for high-fidelity audio)
Video files	<ul style="list-style-type: none">• MPEG-4 (mp4)• JPEG2 2000 (mj2)
Geospatial data	<ul style="list-style-type: none">• Geo-Referenced TIFF (.tiff)• ESRI Shapefile (.shp, .shx, .dbf)• Keyhole Mark Language (.kml)
Software code	<ul style="list-style-type: none">• Plain text (usually with an extension that represents the source language)
Text	<ul style="list-style-type: none">• Plain text (.txt)• Portable Document Format (.pdf)• LaTeX documents (.tex)• Hypertext Markup Language (.html)• Open Document Format (.odt)• Extensible Markup Language (.xml)
E-books	<ul style="list-style-type: none">• EPUB (.epub)

Formats

In which **format** the data will be kept?



Image file:

.tiff -> .jpeg





Audio file:

.wav -> .avi

Video file:

ffv1 -> h264 (.mp4)

Formats

 Text	Preferred	<code>.csv, .txt, .xml</code>
	Acceptable	<code>.csv, .txt, .xml, .rtf, .doc, .xls, .pdf, .ps</code>
 Image	Preferred	<code>.tiff, .pj2</code>
	Acceptable	<code>.tiff, .pj2, .gif, .png, .jpeg</code>
 Audio	Preferred	<code>.wav, .flac</code>
	Acceptable	<code>.aif, .lla, .ogg, .wav, .aac, .mp3, .ram, .wmv</code>
 Video	Preferred	<code>.avi, .mj2, .mpj2</code>
	Acceptable	<code>.avi, .mj2, .mpj2, .mp2, .mp4, .wmv, .mov</code>

II. Description and Organization

Describing and organizing your data makes analysis easier for you and provides context for those you want to share the data with.

Naming Conventions

Consider including:

- Unique Identifiers such as project name or grant number
- Conditions such as lab instrument, solvent..etc
- Date
- Version number

Example1:

aub_ac0000001_000001_m.tif

aub_ac0000001_000001_d.tif

(organization_projectnamebooknumber_page number_master/derivative copy. format)

Example 2:

micro_hhl_20170620_dr_001.tif

(instrument_insectname_date_initials_sequential number.format)



Naming Conventions

System 1: Order by **date**:

2016-04-12_notes_BuildingA.docx

2016-04-12_questionnaire_BuildingA.pdf

2016-12-15_notes_BuildingB.docx

2016-12-15_questionnaire_BuildingB.pdf

System 3: Order by **type**:

Notes_BuildingA_2016-04-12.docx

Notes_BuildingB_2016-12-15.docx

Questionnaire_BuildingA_2016-04-12.pdf

Questionnaire_BuildingB_2016-12-15.pdf

System 2: Order by **subject**:

BuildingA_notes_2016-04-12.docx

BuildingA_questionnaire_2016-04-12.pdf

BuildingB_notes_2016-12-15.docx

BuildingB_questionnaire_2016-12-15.pdf



Documentation and Metadata



What's obvious now might not be in a few months, years, decades...

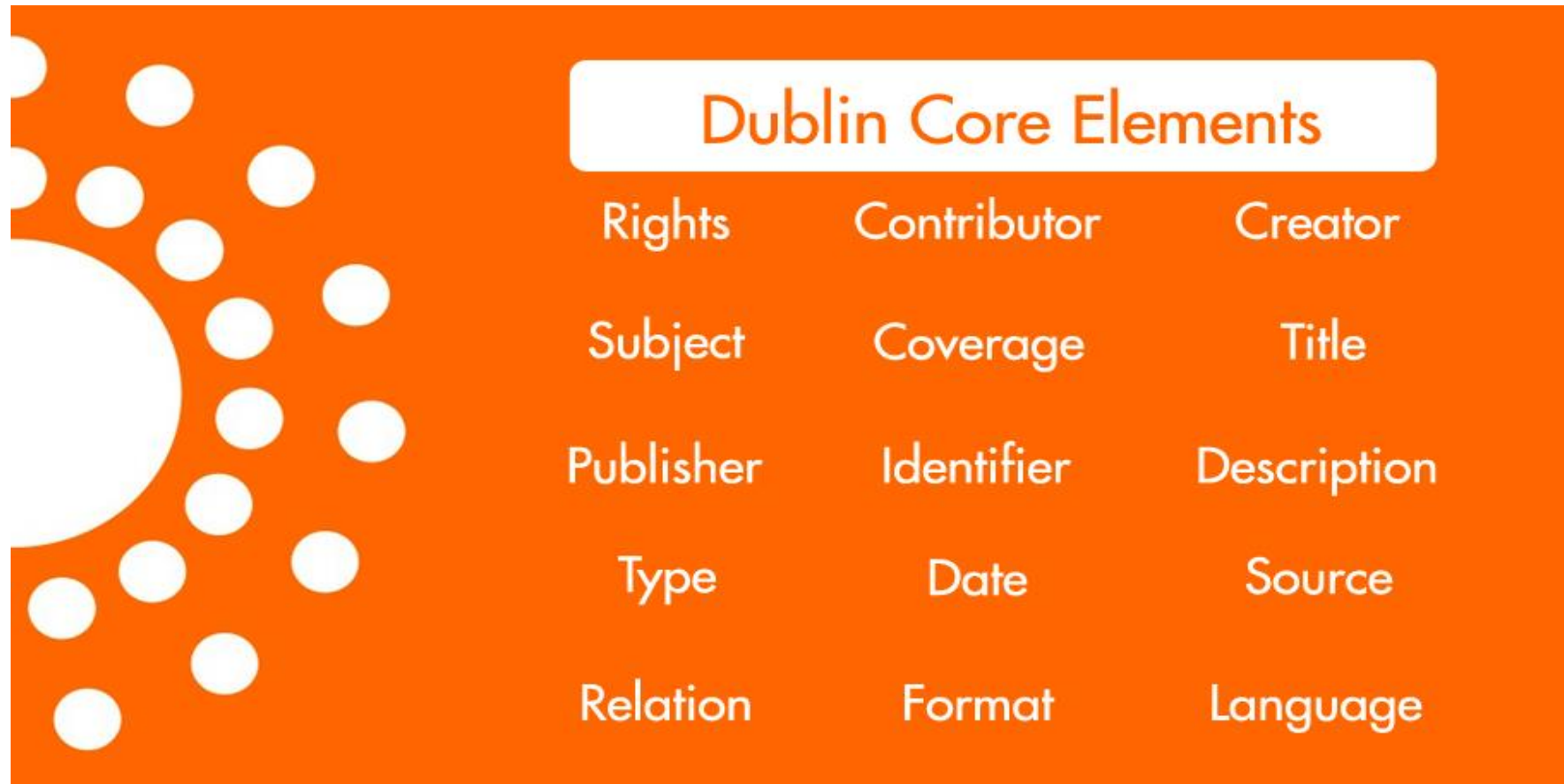


Metadata

“Metadata is structured information that **describes, explains, locates, or otherwise represents** something else.” (NISO, 2004)

- Some fields of research have already defined their metadata standards: Darwin Core (Biology), VRA Core (Visual Art), EBU Core (Audiovisual Content)
- Repositories ask you as a submitter to provide metadata along with your file.

The Dublin Core standard



Documentation



- Detailing your methods helps people understand what you did
- And helps make your work reproducible
- Your conclusions become verifiable

III. Access and Security

How will the data be **backed up**? And how **frequent**?

Who will be **responsible** for the backing up?

How will the data be **recovered** in the event of an incident?

III. Access and Security

Think about where you want to keep several physical and digital copies in case something happens.



III. Access and Security

Consider **AUB campus specific information** on Security:

- IT security: Data Classification Policy
- IRB at AUB (protecting the welfare, rights, and privacy of human subjects)
- AUBMC Access to medical records for research
- AUB Data Bank Institutional Repository Policy

Consider **Laws and Legislations:**

Access to Information Law in Lebanon

HIIPA (Health Insurance Portability and Accountability Act)







Horizon 2020

IV. Reuse and Sharing

- Are you under any obligation to share data?
- How, when, & where will you make the data available?
- What is the process for gaining access to the data?
- Who owns the copyright and/or intellectual property?
- Will you retain rights before opening data to wider use? How long?
- Are permission restrictions necessary?
- Embargo periods for political/commercial/patent reasons?
- Ethical and privacy issues?
- How should your data be cited?

IV. Reuse and Sharing

Creative Commons licenses provide options for selecting how you would like to allow others to use your data.

Icon	Description	Acronym
	Attribution alone	BY
	Attribution + NoDerivatives	BY-ND
	Attribution + ShareAlike	BY-SA
	Attribution + Noncommercial	BY-NC
	Attribution + Noncommercial + NoDerivatives	BY-NC-ND
	Attribution + Noncommercial + ShareAlike	BY-NC-SA

V. Archiving and Preservation

Think about what will happen to your data long-term, beyond its current use in the project.




V. Archiving and Preservation

Deposit your data
in an institutional
repository:

AUB Scholarworks

AUB Data Bank

https://scholarworks.aub.edu.lb/handle/10938/10284



Scholarworks Repository


Scholarworks Home → AUB Data Bank

AUB Data Bank

Browse by

- By Issue Date
- Authors
- Titles
- Subjects

Search within this community and its collections: **Go**



Collections in this community

- [AUB Research Data](#)
This collection includes data generated by AUB faculty members, students and staff.
- [Non-AUB Research Data](#)
This data is collected from private and public institutions.

Recent Submissions

Commercial vs Residential area in Greater Beirut
Center for Remote Sensing; National Council for Scientific Research; gfaour@cnrs.edu.lb (2018-02-07)

معمل الأشغال الفني وإعداد تراخيص المؤسسات السياحية
وزارة السياحة - مصلحة الأبحاث والدراسات والتوثيق: المديرية العامة للتحوزن السياحية
motourism2@gmail.com (2018-01-04)

Total Number of Tourists Arriving to Lebanon by Nationality 2010-2016
Ministry of Tourism; motourism2@gmail.com (2018-01-03)

Road Accidents in Lebanon from 1996-2015
Unknown author (Directorate General of the Internal Security Forces, 2015-11)

Search Scholarworks

Go

Search Scholarworks
 This Community

Browse

All of Scholarworks

- [Communities & Collections](#)
- [By Issue Date](#)
- [Authors](#)
- [Titles](#)
- [Subjects](#)

This Community

- [By Issue Date](#)
- [Authors](#)
- [Titles](#)
- [Subjects](#)

My Account

[Login](#)

Discover

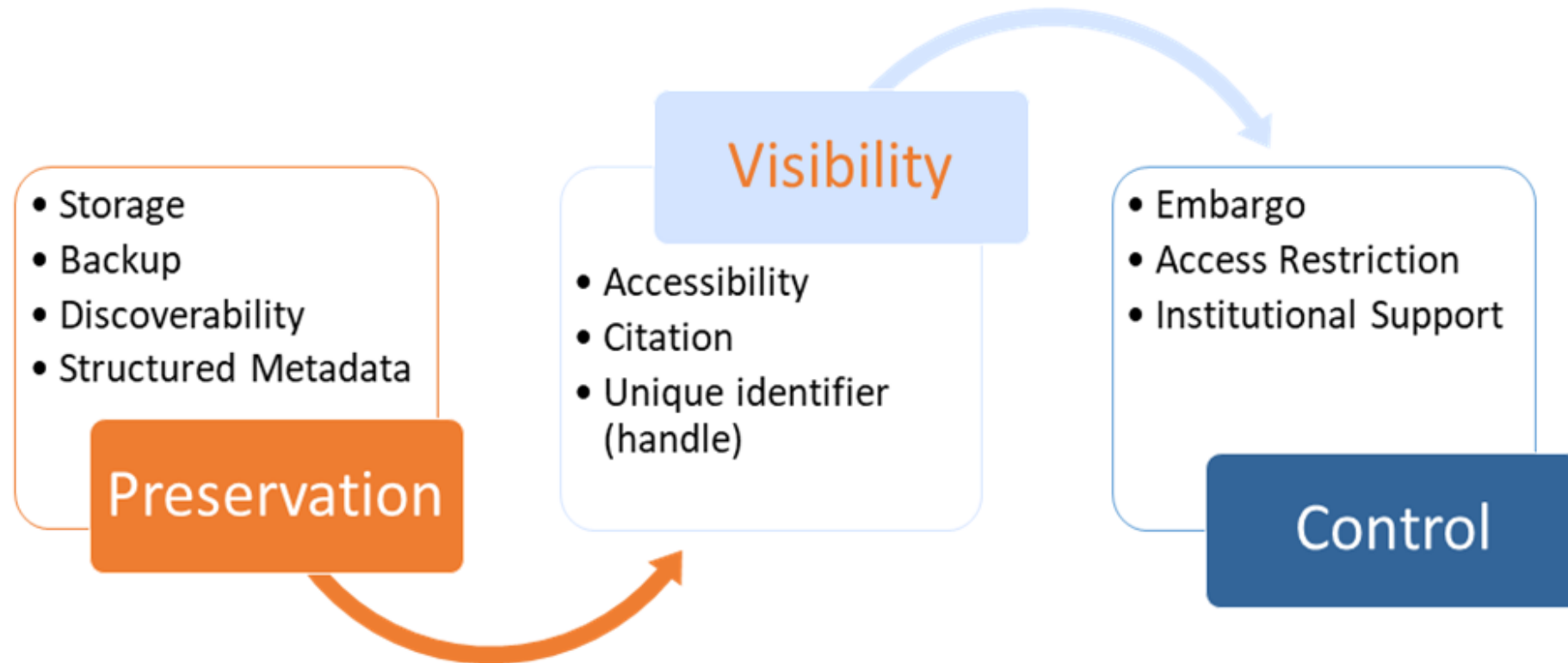
Subject

- [Accidents in Lebanon \(1\)](#)
- [Blood bank \(1\)](#)
- [Blood donation \(1\)](#)
- [Blood donor \(1\)](#)
- [Blood type \(1\)](#)
- [Commercial area Beirut \(1\)](#)
- [Crime \(1\)](#)
- [Death rate \(1\)](#)
- [GIS map Beirut \(1\)](#)
- [Hospitals \(1\)](#)
- [... View More](#)

Date Issued

- [2018 \(3\)](#)
- [2016 \(2\)](#)

AUB Data Bank



V. Archiving and Preservation

There are lots of general and subject specific repositories



Searchable list of repositories by subject or country: <https://www.re3data.org/>

Your turn: RDMP – hands on exercise

- Groups of about four or five people, each focusing on one actual research project
- One members describe their project and data, others act as potential re-user (20 min)
- Write down short notes to the topics of the RDMP and any questions
- Present the plan in 5 minutes

Remember:

- Be brief and simple
- Be specific
- Make your data understandable and reusable

Your Turn:

The DMPTOOL

<https://dmptool.org/>





Welcome

Create data management plans that meet institutional and funder requirements.

Get started

DMPTool by the numbers


29,080
Users


25,417
Plans [More](#)


232
Participating institutions [More](#)

Top 5 templates

- Department of Energy (DOE): Office of Science
- Digital Curation Centre
- NSF-BIO: Biological Sciences
- NSF-SBE: Social, Behavioral, Economic Sciences
- NIH-GEN: Generic

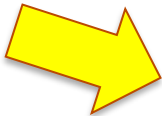
[More](#)

DMPTool News

We have lift off – DMPRoadmap launches!

[Go to the blog](#)





[Learn](#) | [Sign in](#)



- Funder requirements
- Public plans
- Participating institutions
- FAQ
- For researchers
 - Quick start guide
 - Data management general guidance
- For Administrators [↗](#)
 - Promote the DMPTool

Welcome
Create data management plans that meet institutional and funder requirements.

[Get started](#)



Funder guidance
Public plans
Data management resources

DMPTool by the numbers


29,080
Users


25,417
Plans [More](#)


232
Participating institutions [More](#)

Top 5 templates

- Department of Energy (DOE): Office of Science
- Digital Curation Centre
- NSF-BIO: Biological Sciences
- NSF-SBE: Social, Behavioral, Economic Sciences
- NIH-GEN: Generic

[More](#)

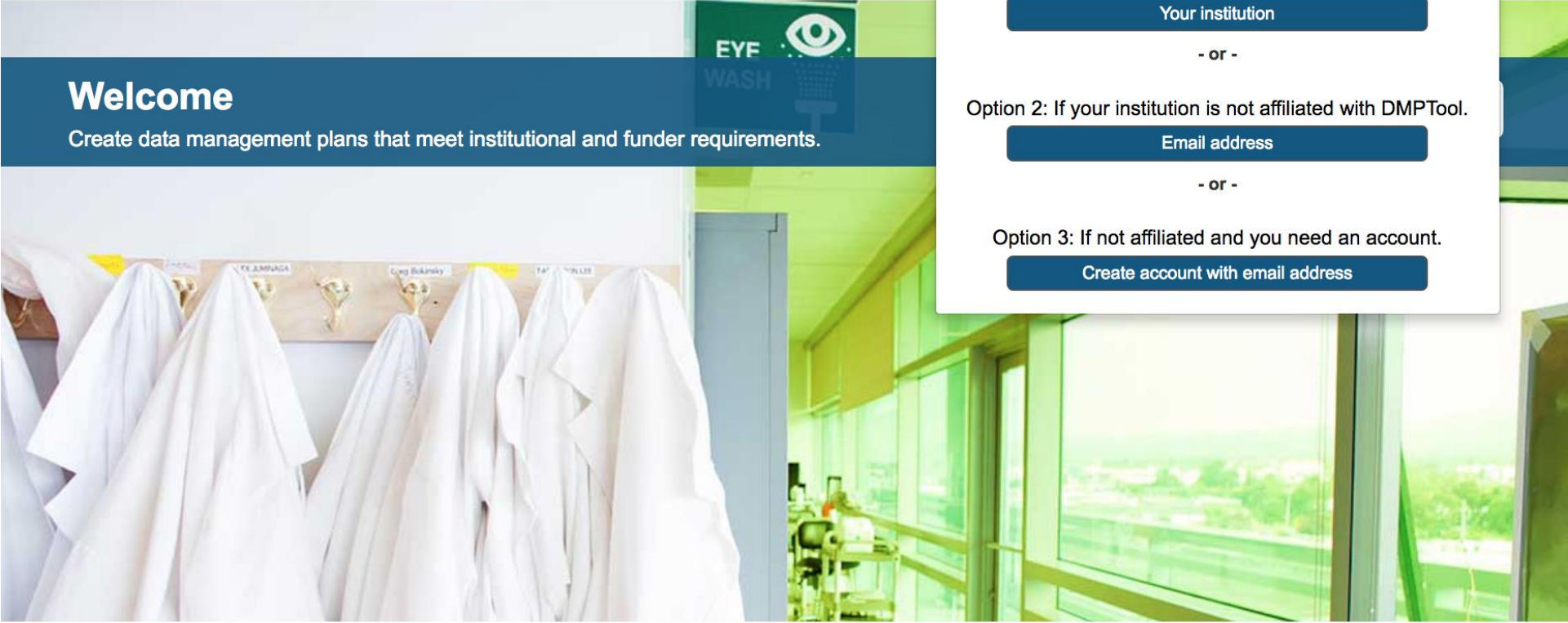
Sign in

[Learn](#) | [Sign in](#)



Welcome

Create data management plans that meet institutional and funder requirements.



Sign in options

Option 1: If your institution is affiliated with DMPTool.

Your institution

- or -

Option 2: If your institution is not affiliated with DMPTool.

Email address

- or -

Option 3: If not affiliated and you need an account.

Create account with email address

DMPTool by the numbers



29,080
Users



25,417
Plans [More](#)



232
Participating institutions [More](#)

Top 5 templates

- Department of Energy (DOE): Office of Science
- Digital Curation Centre
- NSF-BIO: Biological Sciences
- NSF-SBE: Social, Behavioral, Economic Sciences
- NIH-GEN: Generic

[More](#)

Option 1 : Single sign on (for configured institutions)

Look up your institution here ✕

 ✕

Go

See the full list of participating institutions
Institution not in the list? [Create an account with any email address](#)

UC DAVIS
UNIVERSITY OF CALIFORNIA
Central Authentication Service (CAS)

Username:

Passphrase:

LOGIN

OR : Sign in with your institution using an email address
(if not configured for SSO; you must create account first)

Look up your institution here ✕

 ✕

Go

See the full list of participating institutions
Institution not in the list? [Create an account with any email address](#)

UCR ✕

Sign in

* Email

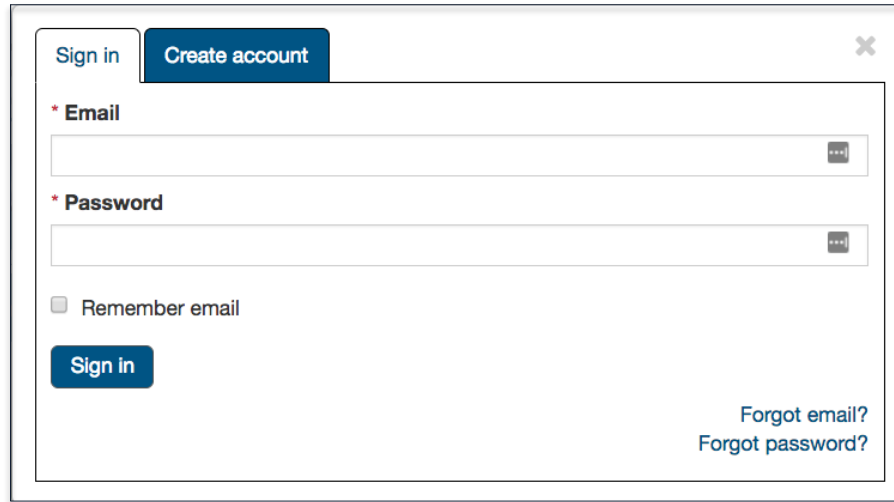
* Password

Remember email

Sign in

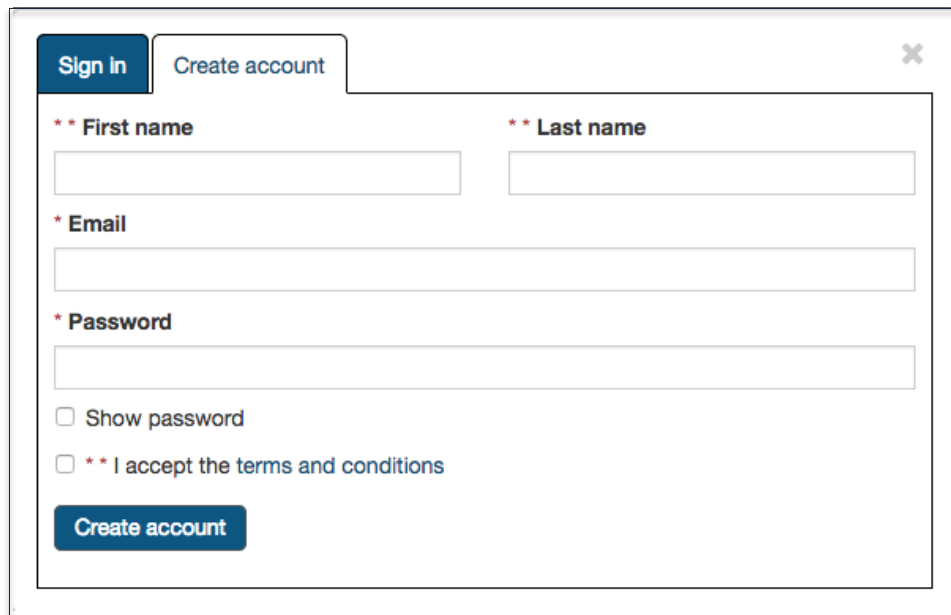
[Forgot email?](#)
[Forgot password?](#)

Option 2 : Sign in as an unaffiliated user



A sign-in form with a modal window. At the top, there are two tabs: "Sign in" (selected) and "Create account". Below the tabs are two input fields: "* Email" and "* Password", each with a clear button (X) on the right. Below the password field is a checkbox labeled "Remember email". At the bottom left is a blue "Sign in" button. At the bottom right are two links: "Forgot email?" and "Forgot password?".

Option 3 : Create an account with an email address



A create account form with a modal window. At the top, there are two tabs: "Sign In" and "Create account" (selected). Below the tabs are four input fields: "** First name", "** Last name", "* Email", and "* Password". Below the password field is a checkbox labeled "Show password" and another checkbox labeled "** I accept the terms and conditions". At the bottom is a blue "Create account" button.

Edit profile

Personal details

Password

Notification preferences

Please note that your email address is also your username. If you change this remember to use your new email address on sign in. If your account is created with your institutional credentials you must contact us to change your email or organization.

* Email

testuser@gmail.com

* First name

Test

* Last name

User

Organization ?

University of California, Office of the President (UCOP)

ORCID

 Create or connect your ORCID ID

Save

Set up your
profile and
notification
preferences

Add your ORCID

 Create a plan

My dashboard

[Create plan](#)

The table below lists the plans that you have created, and that have been shared with you by others. You can edit, share, download, make a copy, or remove these plans at any time.

Project title ▾	Template ▾	Last edited ^	Permissions	Test	Visibility	Shared	Actions ▾
Prueba123	NIH-GEN: Generic	03-14-2018	Owner	<input checked="" type="checkbox"/>	N/A	No	
DMPRoadmap: Making Data Management Plans Actionable	NSF-GEN: Generic	03-18-2018	Co-owner	<input type="checkbox"/>	Public	Yes	
A unified approach to preserving cultural software objects a ...	NEH-ODH: Office of Digital Humanities	03-29-2018	Owner	<input type="checkbox"/>	Private	Yes	

- Edit
- Share
- Download
- Make a copy
- Remove

University of California, Office of the President (UCOP) Plans

The table below lists the plans that users at your organization have created and shared within your organization. This allows you to download a PDF and view their plans as s to discover new research data.

Project title ^	Template ▾	Owner	Last edited ▾	Download
A unified approach to preserving cultura ...	NEH-ODH: Office of Digital Humanities	dmpcurator@gmail.com	06-05-2014	PDF
Arthropod responses to grassland nutrien ...	NSF-GEN: Generic	dmpcurator@gmail.com	05-30-2014	PDF
Atmospheric CO2 Concentrations, Mauna Lo ...	NSF-AGS: Atmospheric and Geospace Sciences	dmpcurator@gmail.com	05-30-2014	PDF

My Dashboard:

1. My plans
2. Plans created by others at my organization & shared internally

Create a new plan

Before you get started, we need some information about your research project to set you up with the app.

What research project are you planning?

Arthropod responses to grassland nutrient limitation



Mock project for testing, practice, or educational purposes

Select the primary research organization

University of California, Office of the President (UCOP)



- or -

My research organization is not on the list or no research organization is associated with this plan

Select the primary funding organization

National Science Foundation (NSF)



- or -

No funder associated with this plan

Which template would you like to use?

NSF-GEN: Generic



We found multiple templates corresponding to your funder.

Create plan

Cancel

Check box below for
test or practice plans

1. Enter project title
2. Select research organization
3. Select funder
4. (If a funder has multiple templates)
Select template

Arthropod responses to grassland nutrient limitation

Project details

Plan overview

Write plan

Share

Download

* Project title

Arthropod responses to grassland nutrient limitation

mock project for testing, practice, or educational purposes

Funder

National Science Foundation (NSF)

Grant number

Project abstract

B *I*

Principal Investigator

Name

Test User

ORCID iD

Email

stephanie.simms@gmail.com

Data contact person

Same as Principal Investigator

Submit

Cancel

Plan guidance configuration

To help you write your plan, DMPTool can show you guidance from a variety of organizations.

Select up to 6 organizations to see their guidance.

- DMPTool
- University of California, Riverside (UCR)

Find guidance from additional organizations below

[See the full list](#)

Submit

Add an abstract, other info about the project & select guidance

Arthropod responses to grassland nutrient limitation

[Project details](#)[Plan overview](#)[Write plan](#)[Share](#)[Download](#)

NSF-GEN: Generic

This plan is based on the "NSF-GEN: Generic" template provided by National Science Foundation (NSF).

Instructions

[Write plan](#)

Types of data produced

- The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project.

Data and metadata standards

- The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies).

Policies for access and sharing

- Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

Policies for re-use, re-distribution, derivatives

- Policies and provisions for re-use, re-distribution, and the production of derivatives.

Plans for archiving and preservation

- Plans for archiving data, samples, and other research products, and for preservation of access to them.

Get an overview of the questions
the funder will ask you to address

Arthropod responses to grassland nutrient limitation

Project details | Plan overview | Write plan | Share | Download

expand all | collapse all 0/5 answered

Types of data produced (0 / 1)

The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project.

B I [List icons] [Link icon] [Table icon]

Save

Guidance | **Comments**

NSF | **DMPTool**

Guidance

- NSF Proposal & Award Policies & Procedures Guide (PAPPG)
- NSF plans for data management and sharing of the products of research (PAPPG)
- NSF Dissemination and Sharing of Research Results
- NSF Frequently Asked Questions (FAQs) for Public Access

+ Data and metadata standards (0 / 1)

+ Policies for access and sharing (0 / 1)

+ Policies for re-use, re-distribution, derivatives (0 / 1)

Get instructions and guidance; leave comments for collaborators

Open each accordion to answer the questions

Arthropod responses to grassland nutrient limitation

[Project details](#)[Plan overview](#)[Write plan](#)[Share](#)[Download](#)

Set plan visibility

Public or organizational visibility is intended for finished plans. You must answer at least 50% of the questions to enable these options. Note: test plans are set to private visibility by default.

- Private: visible to me, specified collaborators and administrators at my organization
- Organizational: anyone at my organization can view
- Public: anyone can view

Share your plan
(optional)

Manage collaborators

Invite specific people to read, edit, or administer your plan. Invitees will receive an email notification that they have access to this plan.

Email address	Permissions	
stephanie.simms@gmail.com	Owner	
brian.riley@ucop.edu	Co-owner ↕	Remove

Invite collaborators

* Email

* Permissions

- Co-owner: can edit project details, change visibility, and add collaborators
- Editor: can comment and make changes
- Read only: can view and comment, but not make changes

Send invitation

Invite collaborators to
coauthor the plan

Request expert feedback

Click below to give data management staff at your organization access to read and co

You can continue to edit and download the plan in the interim.

Request feedback

Request feedback (if enabled
for your organization)

Arthropod responses to grassland nutrient limitation

Project details

Plan overview

Write plan

Share

Download

Download settings

Optional plan components

- project details coversheet
- question text and section headings
- unanswered questions

Format

docx

Download plan

Download your plan
(docx, pdf, txt, html, csv)

DMPTool Blog

Guidance & resources for
your data management plan



[About the DMPTool](#)

[Go to the Tool](#)

[DMPTool Webinars](#)

[Presentations and Press](#)

[Maintenance](#)

Posted on **February 27, 2018** by [stephaniesimms](#)

[← Previous](#)

[Next →](#)

New DMPTool launched today!



We're delighted to announce a successful launch of DMPTool version 3 today. This milestone represents the convergence of the two most popular data management planning tools—US-based DMPTool and UK-based [DMPonline](#)—into a single, internationalized platform. We plan to bring the many other installations of the tool in Canada, Australia, South Africa, Argentina, and throughout Europe along for the ride as we work together to make DMPs a more useful exercise for everyone!

Currently the DMPTool supports 226 institutions and more than 28,000 users worldwide. The new DMPTool retains all of the existing functionality plus some handy new things, all in a shiny new package:

For everyone

- Up-to-date [Funder requirements](#) (34 templates and more to come...)
- A searchable list of [participating institutions](#) with & without single sign-on
- New user interface with streamlined plan writing pages
- Website translations coming soon (Spanish, French, German, Brazilian Portuguese, Japanese, [contact us](#) to contribute another language!)
- [Quick-start guide for creating a DMP](#) (downloadable pdf coming soon)
- [General data management guidance](#)

Thank you