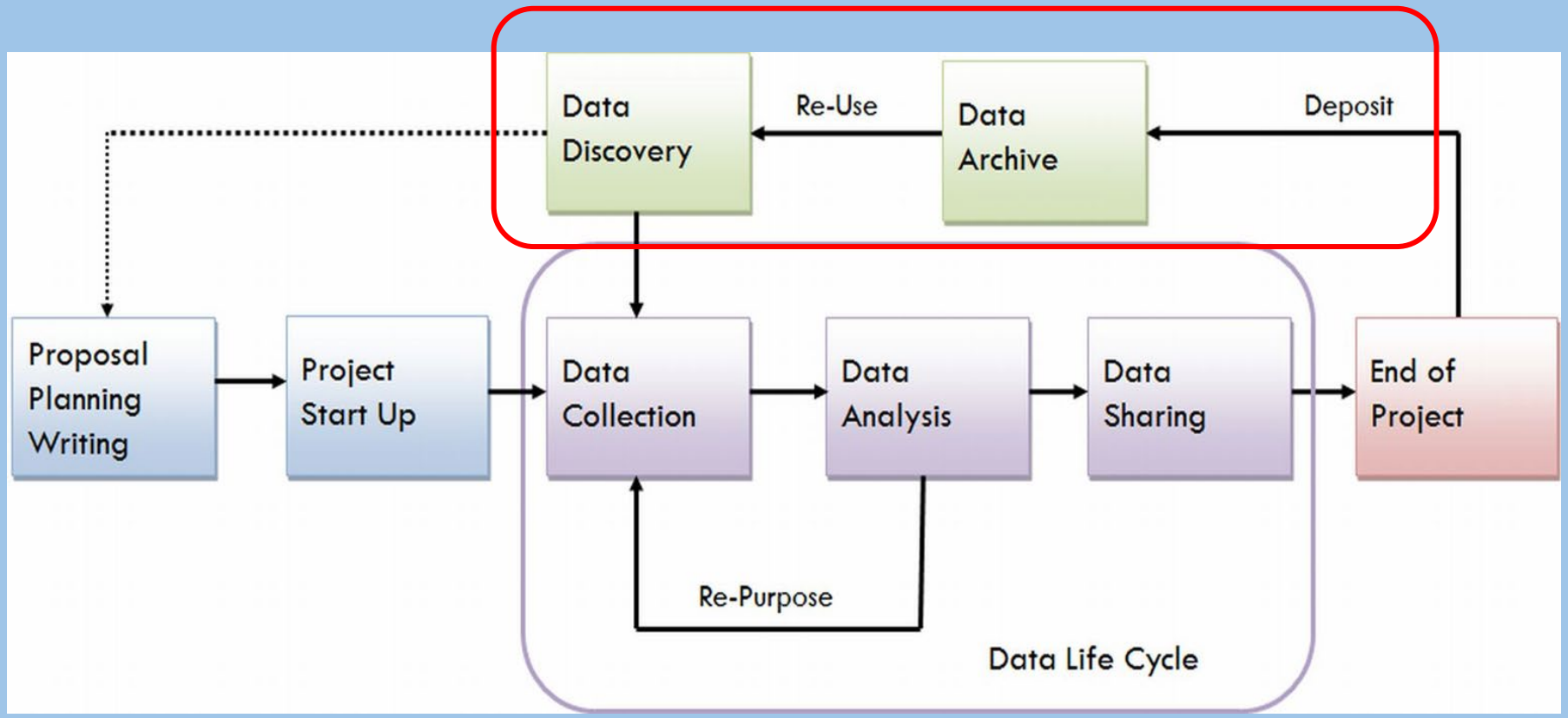
Support your research with data management planning!

James Doiron
RDM Services Coordinator, University of Alberta Library
Academic Director, UofA Research Data Centre
Co-chair, Portage DMP Expert Group

March 23, 2021

What is Research Data Management?

RDM = managing data throughout all phases of the research lifecycle….through active phases and beyond

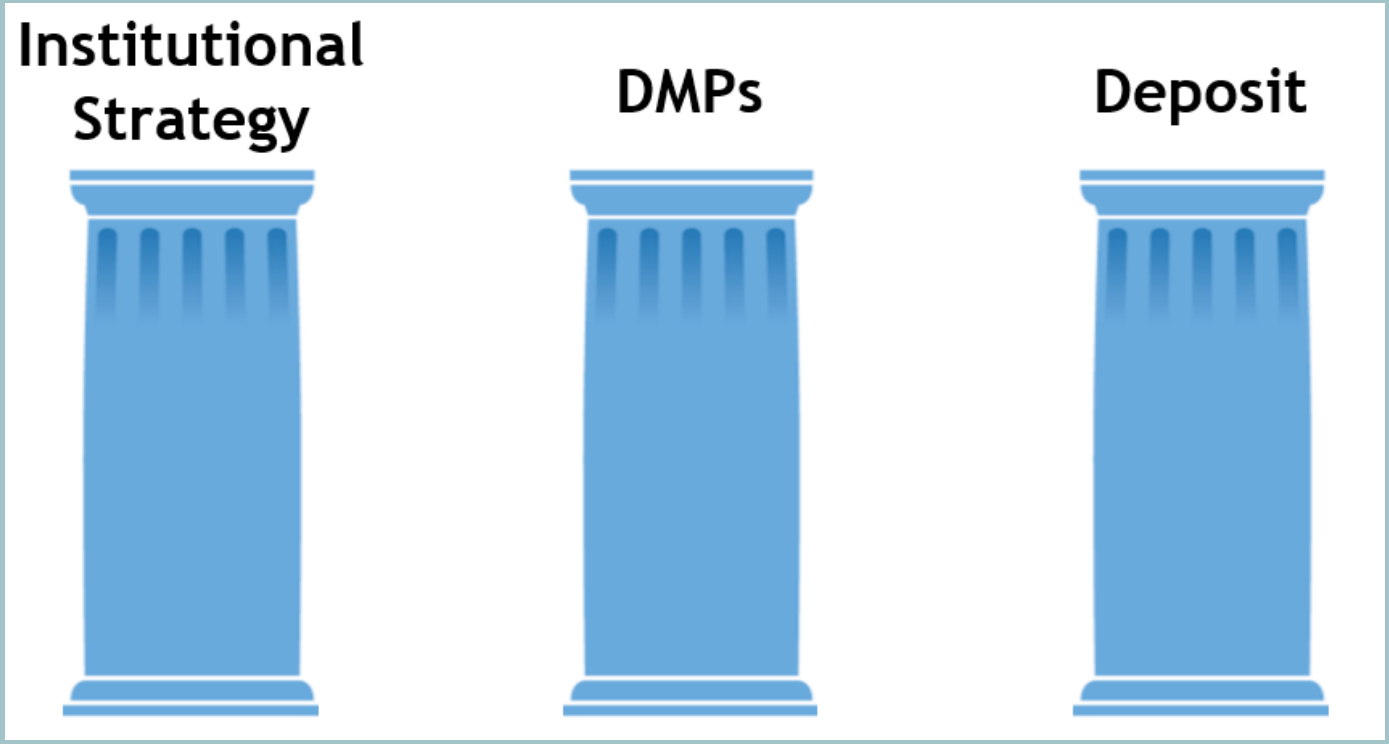


Source: " *E- Science and the Life Cycle of Research*" , Humphrey, C. (2006)

Tri-Agency RDM Policy

In March 2021, the Tri-Agency released their
Research Data Management Policy

The policy includes general requirements and implementation dates across three main pillars:



Tri-Agency RDM Policy - “Take away points”

Institutional Strategy:

Each postsecondary institution eligible to administer Tri-Agency funds is required to create and release an institutional RDM strategy by March 2023

Data Management Plans:

For certain funding opportunities, data management plans (DMPs) will be required to be submitted at the time of application, as outlined in the call for proposals; in these cases, the DMPs will be considered in the adjudication process.

NOTE: By spring 2022, the agencies will identify the initial set of funding opportunities subject to the DMP requirement. The agencies will pilot the DMP requirement in targeted funding opportunities before this date.

Data Deposit:

Grant recipients are required to deposit into a digital repository all digital research data, metadata and code that directly support the research conclusions in journal publications and pre-prints that arise from agency-supported research.

NOTE: After reviewing the institutional strategies and in line with the readiness of the Canadian research community, the agencies will phase in the deposit requirement.

FAIR https://www.go-fair.org/fair-principles/

FAIR is a set of guiding principles focused towards making data:

- **Findable** - Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.
- **Accessible** - Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.
- **Interoperable** - Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- **Reusable** - Data and collections have clear usage licenses and provide accurate information on provenance.

*Key Reading:

Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/ sdata.2016.18 (2016).

FAIR Principles: Key Resource

go-fair.org (https://www.go-fair.org/fair-principles/)

- Detailed information across the FAIR principles
- Implementation Networks
- News
- Event
- Resources!

OCAP

The First Nation Principles of OCAP are a set of standards that establish how First Nations data should be collected, protected, used or shared:

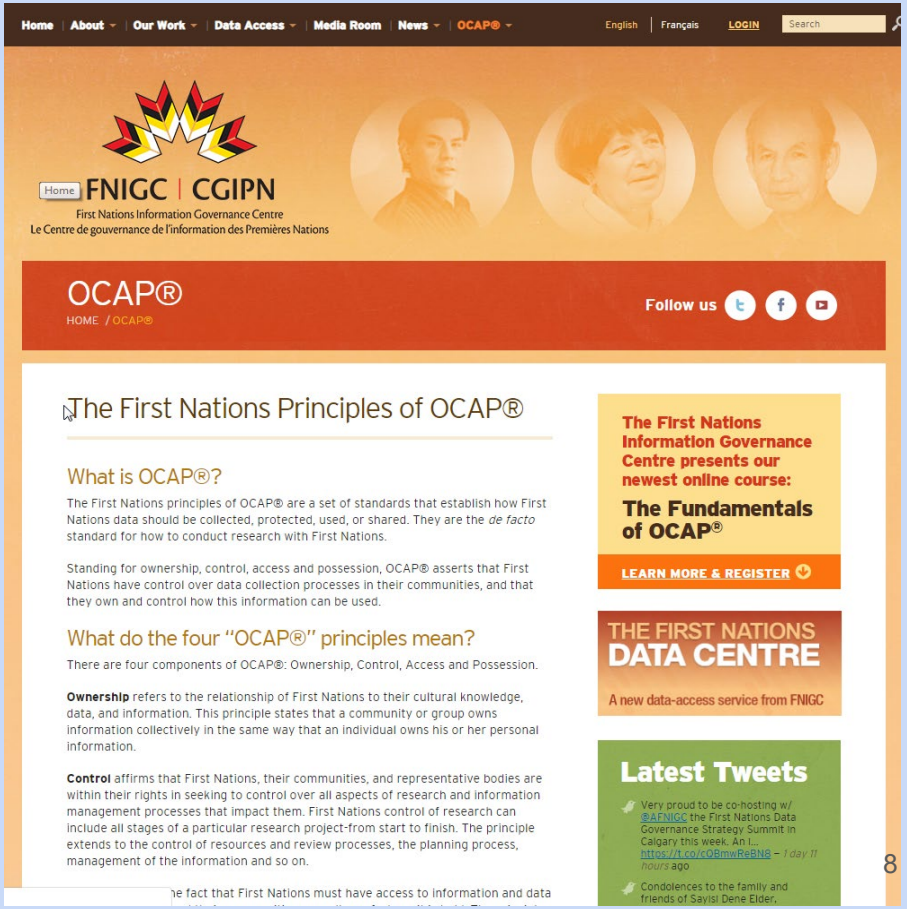
- **Ownership:** refers to the relationship of First Nations to their cultural knowledge, data & information - a **community/group collectively owns information** in the same way that an individual owns his/her personal information
- **Control:** affirms that **First Nations communities have rights in seeking control over all aspects of research** - from start to finish - that impact them. This extends to control of resources and review processes and management of information.
- **Access: First Nations must have access to information and data about themselves and their communities** regardless of where it is held, and have the **right to manage and make decisions** regarding access to their collective information.
- **Possession:** Refers to the physical control of data - the **mechanism by which ownership can be asserted and protected.**

"OCAP® is a registered trademark of the First Nations Information Governance Centre (FNIGC)"
www.FNIGC.ca/OCAP

OCAP Principles: Key Resource

First Nations Information Governance Centre
(https://www.FNIGC.ca/)

- Fundamentals of OCAP online training program
- FNIGC data online
- First Nations Data Centre (data by request)
- First Nations surveys (i.e., regional health, early childhood, education, labour, oral health)
- FNIGC online library



What *is* a data management plan (DMP)?

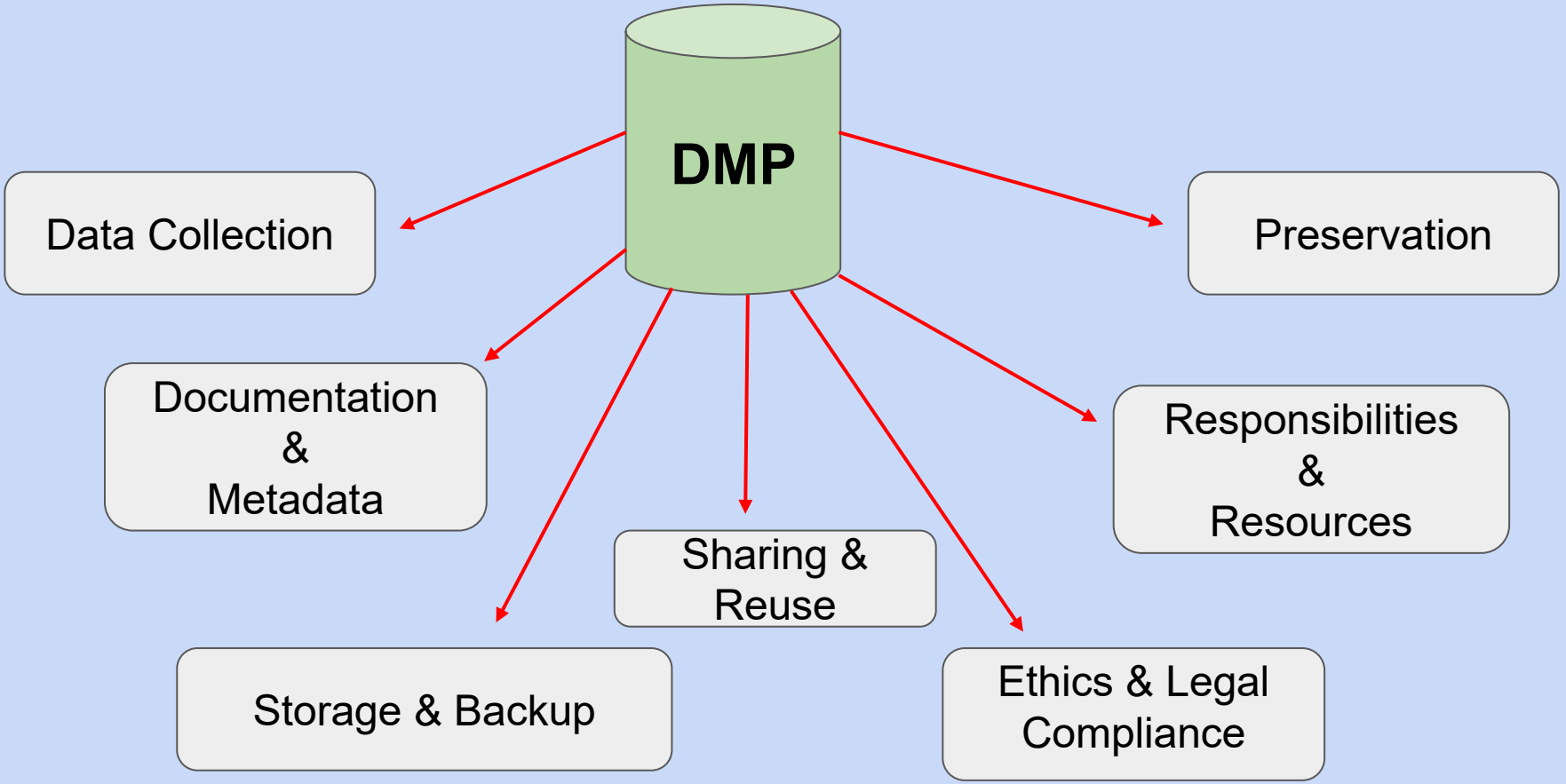
A DMP:

- Is a formal document which clearly articulates the strategies and tools you will implement to effectively manage your data

- Speaks to the management of data both ***during*** the active phases of your research and ***after*** the completion of the research project.

The objective of a DMP is to address issues related to data management prior to starting your research project!

A DMP provides information across key research lifecycle categories:



Why are DMPs important?

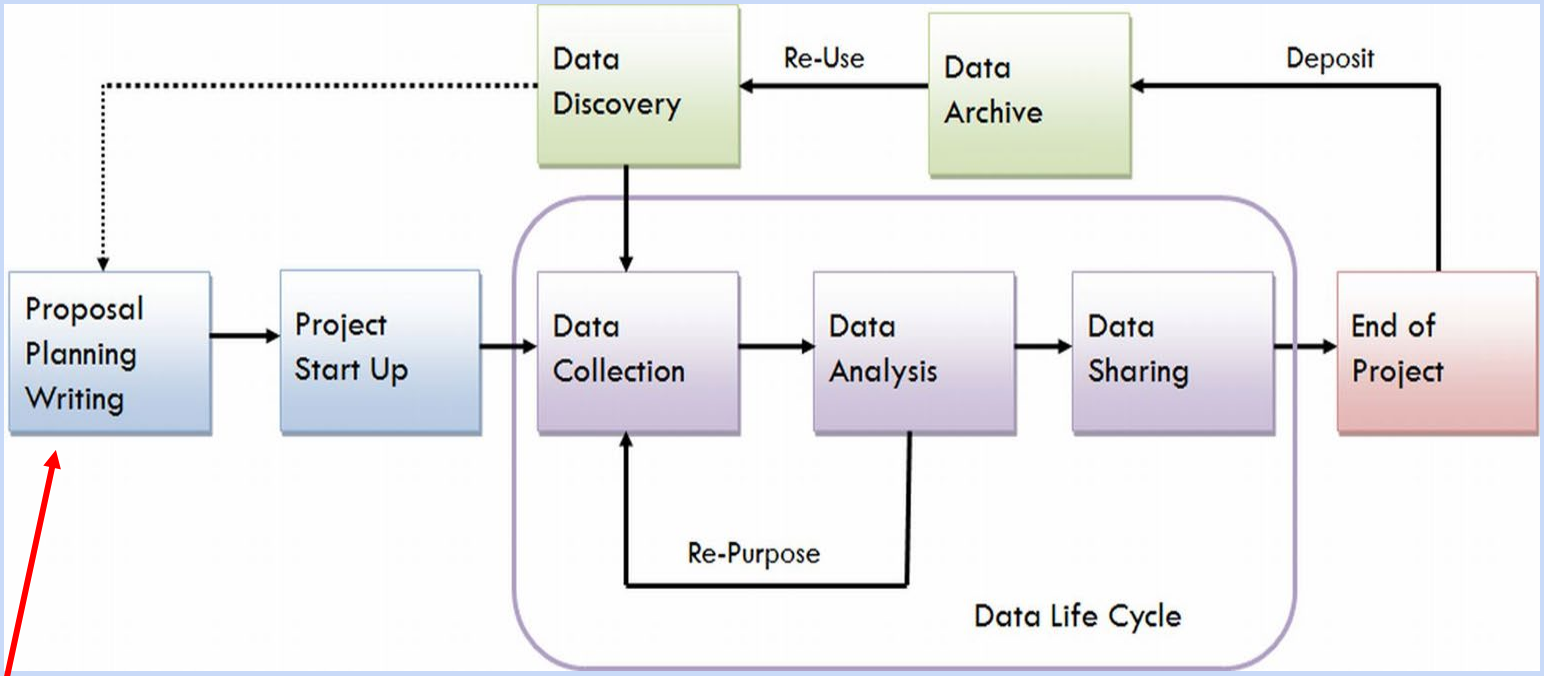
A data management plan is important to the research process as it can help you to:

- set out consistent strategies prior to starting your research for how data will be managed throughout its entire lifecycle
- identify the strengths & weaknesses in your current practices and make decisions on how to integrate effective data management practices into your process
- prepare data for future reuse, preservation and sharing
- reduce the overall cost of research by increasing project efficiencies

When should I start creating a DMP?

A DMP should be developed as early in the research process as possible!

Every research project should ***begin*** with the creation of a DMP, as it will be used to guide the research process.



DMP

DMP General guidelines

Begin by providing a description of your research project, its focus, and purpose

Avoid extensive use of discipline specific jargon - your DMP should be easily understood by anyone!

Provide clarification for any acronyms used

Do not leave sections or questions blank

Provide rationale for decisions made - help others understand *why* you have made a decision

Your DMP is a living document - update it as needed!

DMP Sections: Data Collection

Include descriptions of *how* you will collect data, including from where and in what format(s)

Provide an estimate of the amount of data you will collect (e.g., MBs/GBs/TBs).

Describe any software and/or platforms that will be used for data collection

Explain how you will organize your data, including details relating both to file naming and versioning

Clearly explain how you will both store and transfer data

DMP Sections: Documentation & Metadata

Describe what information will be needed for others to understand or reuse your data.

Describe how you will consistently capture documentation throughout the project.

Choose a metadata standard suited to your discipline and/or chosen data repository, or provide rationale for creating your own.

DMP Sections: Storage & Backup

Provide an estimate of storage space needed during the active phases of your research - remember to take into account file versioning, backups, and data growth!

If needed, follow the 3-2-1 backup rule: 3 copies of your data, on 2 different storage media, with 1 located offsite.

State a data backup schedule, automatic being most ideal.

Describe how collaborators or research team will be able to access, modify, contribute, and work with your data.

DMP Sections: Preservation

Not all data that you create necessarily *needs* to be preserved - consider such things as the value of your data, funding requirements, etc., and decide which, if any, should be preserved. Consult with experts in the Libraries as needed!

Consider optimal file formats (preferably non-proprietary) for supporting long-term preservation. Optimally preserved data are easily accessible and use by anyone, without requiring proprietary software to do so.

DMP Sections: Sharing & Reuse

Consider the appropriate sharing of your data, including any funding or confidentiality requirements.

Explain what uses can be made of your data through licenses like Creative Commons.

Consult with colleagues or librarians to choose an appropriate data repository or search re3data.org to find one.

Choose a repository that assigns permanent identifiers to datasets (e.g., DOI) to enhance discoverability, accessibility, and citability.

If applicable, describe how you will ensure file integrity, anonymization and de-identification.

DMP Sections: Responsibilities & Resources

Identify data stewardship roles and responsibilities of project members and other organizations during and after the project.

Estimate and describe any required resources and costs for data management and long-term access to your data.

DMP Sections: Ethics & Legal Compliance

Describe if there are any legal, ethical, and intellectual property issues when managing and sharing your data.

Explain how you will comply with any applicable privacy legislation and laws, including funding and institutional requirements.

Describe how you will ensure your data are securely managed after the project is completed including the secure management of sensitive data and in accordance with any ethical obligations

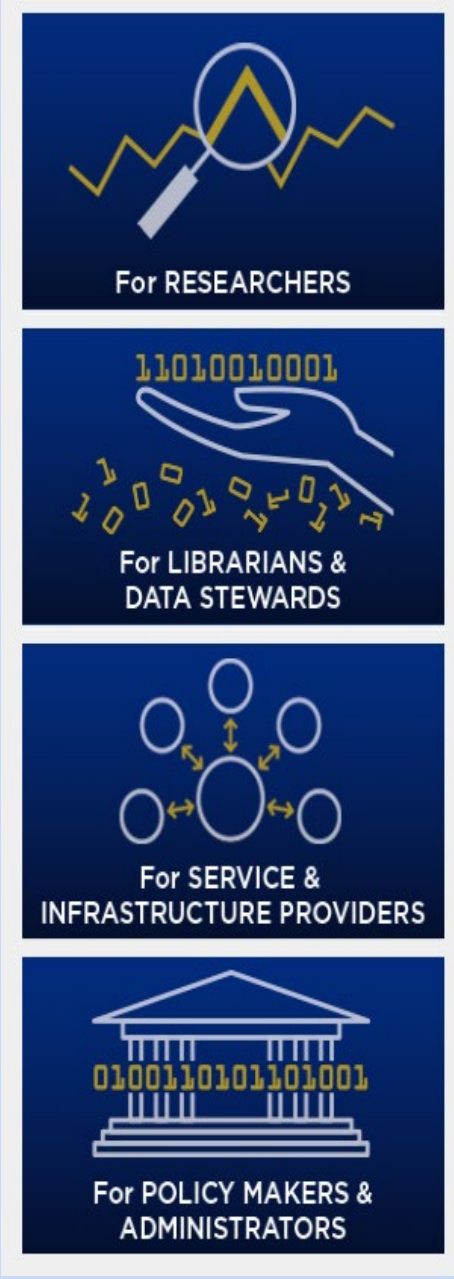
Portage Network

Portage is a national RDM network launched by CARL in 2015 which coalesces initiatives to build capacity and coordinate RDM activities in Canada

1. **Networks of Expertise:**
- Pan-Canadian RDM expertise
- Provides resources, tools, and experts in the area of RDM

1. **Infrastructure Platforms**
- Works to assemble and deliver essential RDM infrastructure and service components

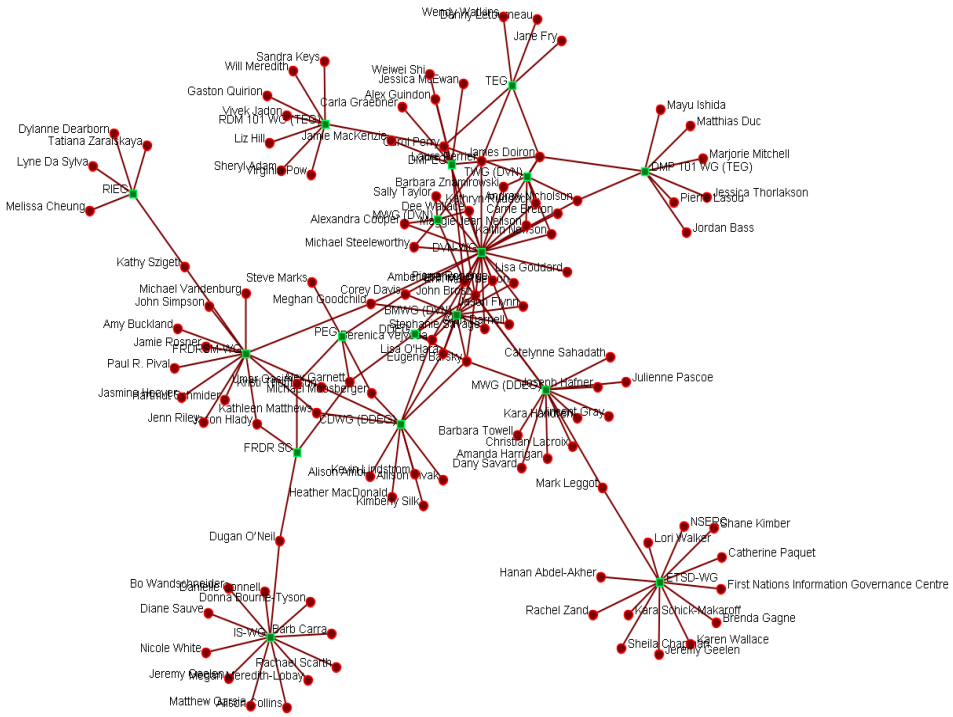
**Portage provides bi-lingual expertise and resources*



Based on a national network of experts

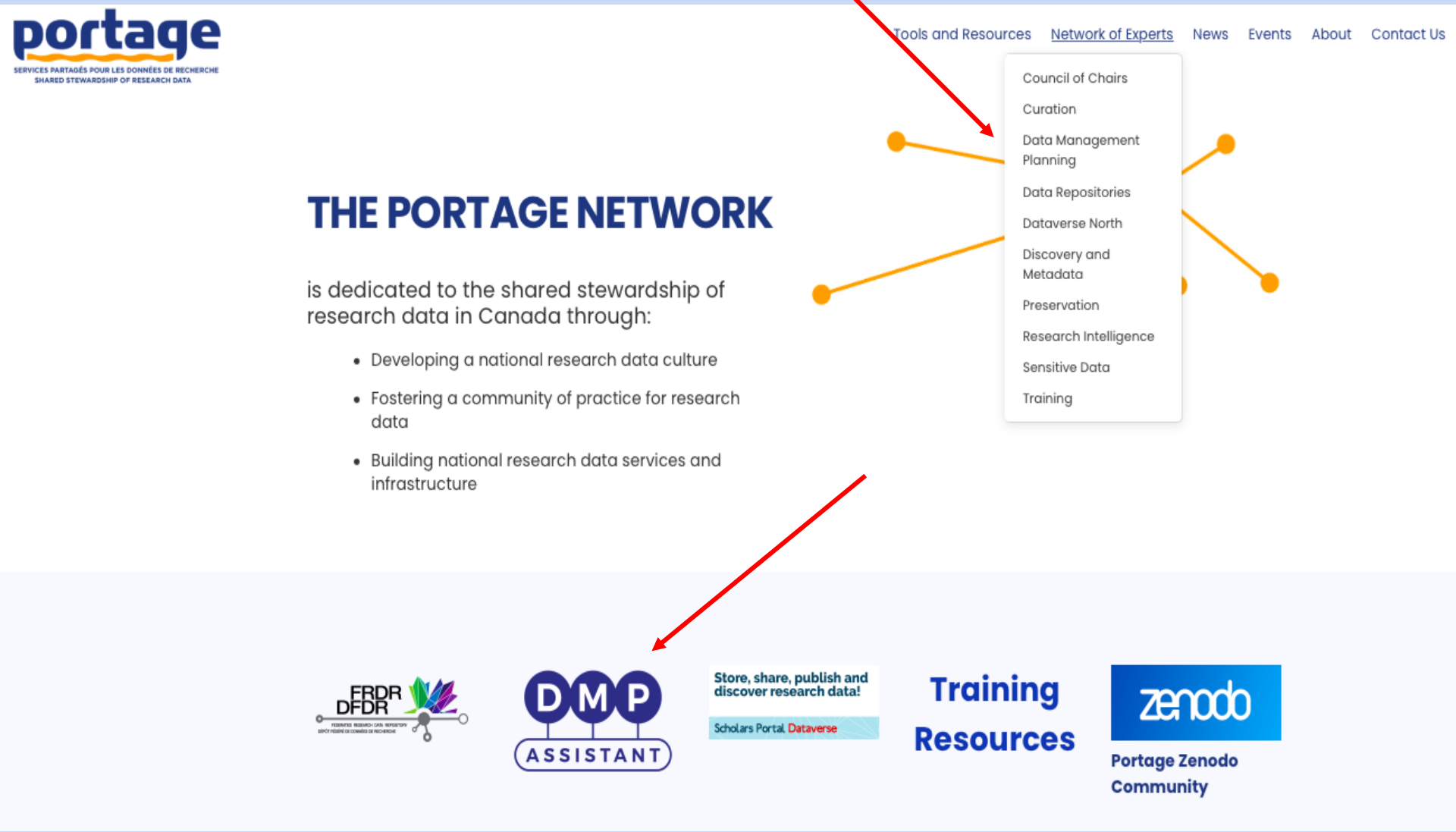
Expert Groups

- **Data Management Planning**
- Curation
- Data Discovery & Metadata
- Preservation
- RDM Training
- Research Intelligence
- Data Repositories
- Sensitive Data
- Dataverse North WG

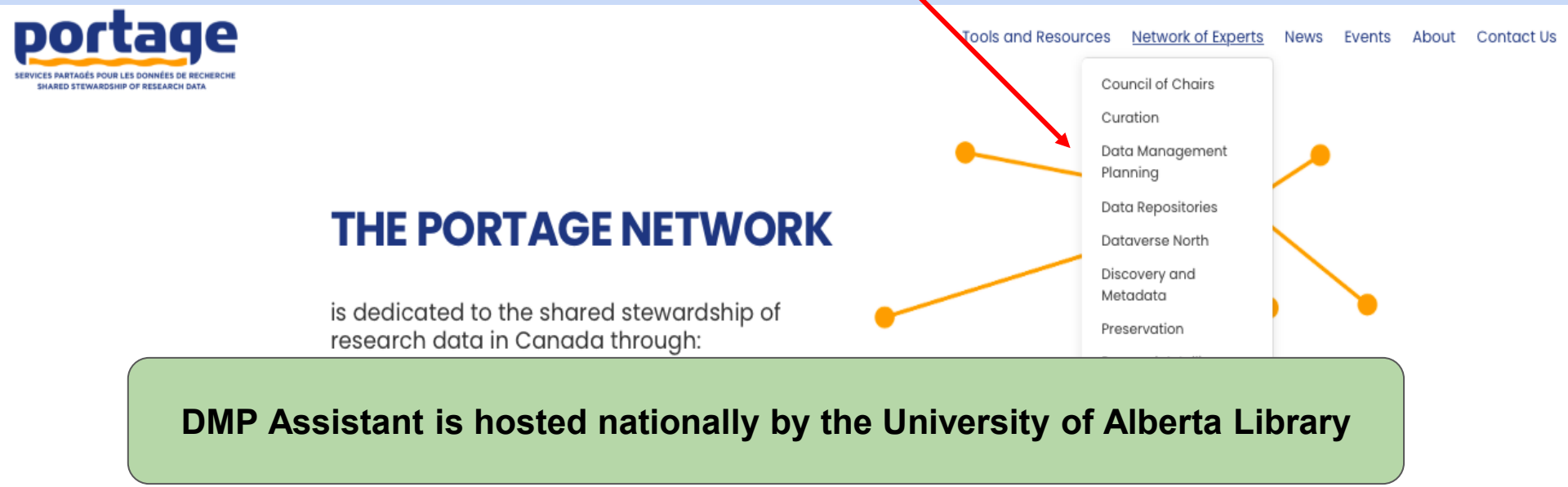


portage

portagenetwork.ca



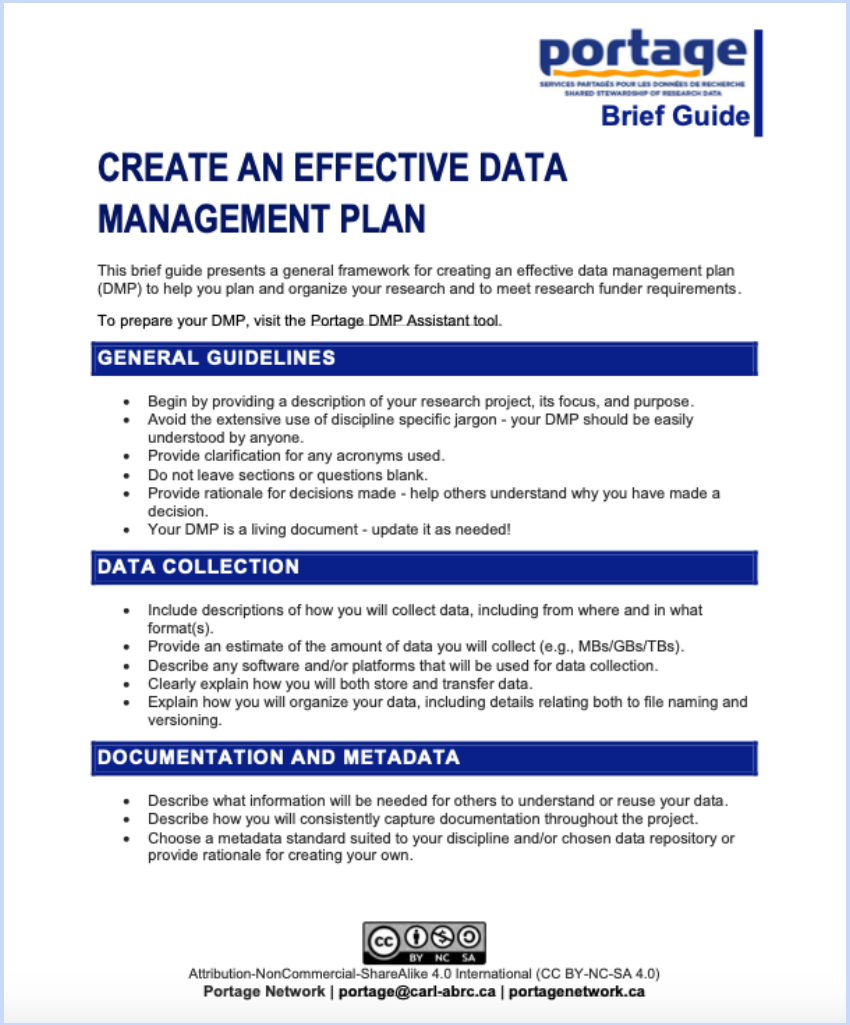
portagenetwork.ca



DMP Assistant is hosted nationally by the University of Alberta Library

DMP guidance resources

Creating an effective DMP (English/French)



portage
SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE
SHARED STEWARDSHIP OF RESEARCH DATA
Brief Guide

CREATE AN EFFECTIVE DATA MANAGEMENT PLAN

This brief guide presents a general framework for creating an effective data management plan (DMP) to help you plan and organize your research and to meet research funder requirements.

To prepare your DMP, visit the Portage DMP Assistant tool.

GENERAL GUIDELINES

- Begin by providing a description of your research project, its focus, and purpose.
- Avoid the extensive use of discipline specific jargon - your DMP should be easily understood by anyone.
- Provide clarification for any acronyms used.
- Do not leave sections or questions blank.
- Provide rationale for decisions made - help others understand why you have made a decision.
- Your DMP is a living document - update it as needed!

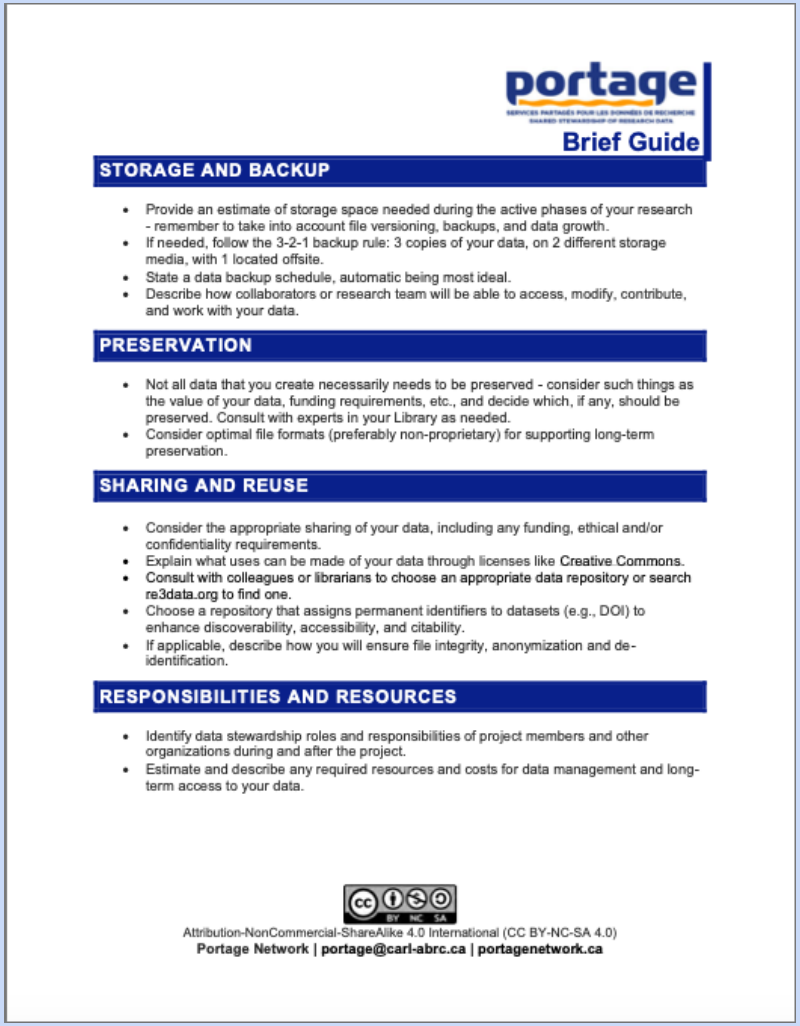
DATA COLLECTION

- Include descriptions of how you will collect data, including from where and in what format(s).
- Provide an estimate of the amount of data you will collect (e.g., MBs/GBs/TBs).
- Describe any software and/or platforms that will be used for data collection.
- Clearly explain how you will both store and transfer data.
- Explain how you will organize your data, including details relating both to file naming and versioning.

DOCUMENTATION AND METADATA

- Describe what information will be needed for others to understand or reuse your data.
- Describe how you will consistently capture documentation throughout the project.
- Choose a metadata standard suited to your discipline and/or chosen data repository or provide rationale for creating your own.

Portage Network | portage@carl-abrc.ca | portagenetwork.ca



portage
SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE
SHARED STEWARDSHIP OF RESEARCH DATA
Brief Guide

STORAGE AND BACKUP

- Provide an estimate of storage space needed during the active phases of your research - remember to take into account file versioning, backups, and data growth.
- If needed, follow the 3-2-1 backup rule: 3 copies of your data, on 2 different storage media, with 1 located offsite.
- State a data backup schedule, automatic being most ideal.
- Describe how collaborators or research team will be able to access, modify, contribute, and work with your data.

PRESERVATION

- Not all data that you create necessarily needs to be preserved - consider such things as the value of your data, funding requirements, etc., and decide which, if any, should be preserved. Consult with experts in your Library as needed.
- Consider optimal file formats (preferably non-proprietary) for supporting long-term preservation.

SHARING AND REUSE

- Consider the appropriate sharing of your data, including any funding, ethical and/or confidentiality requirements.
- Explain what uses can be made of your data through licenses like Creative Commons.
- Consult with colleagues or librarians to choose an appropriate data repository or search re3data.org to find one.
- Choose a repository that assigns permanent identifiers to datasets (e.g., DOI) to enhance discoverability, accessibility, and citability.
- If applicable, describe how you will ensure file integrity, anonymization and de-identification.

RESPONSIBILITIES AND RESOURCES

- Identify data stewardship roles and responsibilities of project members and other organizations during and after the project.
- Estimate and describe any required resources and costs for data management and long-term access to your data.

Portage Network | portage@carl-abrc.ca | portagenetwork.ca

DMP exemplars

Discipline and/or methodology focus	Title & Link
Digital Humanities	Belgians and French in the Prairies
	Soundscape Study
	Women's Print History Project (1750-1830)
	Historical Canadian Census Data
Mixed Methods (surveys & qualitative interviews)	Mixed Methods Fictional Exemplar
Natural Sciences	Ecohydrology Research Group
	Computational Reproducibility in High-Performance Computing
Social Sciences	People, Places, Policies & Prospects: Affordable Rental Housing for Those in Greatest Need
	Usage of Academic Profile Websites

DMP exemplars

Discipline and/or methodology
Digital Humanities
Mixed Methods (surveys & qualitative interviews
Natural Sciences
Social Sciences

portage

Data Management Plan Exemplar #3: Mixed Methods

Fictional Exemplar

Data Collection

What types of data will you collect, create, acquire and/or record?

We will be collecting surveys which will then be exported into tabular format.

We will also be conducting both semi-structured interviews and focus groups that will produce both digital audio and text (transcriptions) based data.

What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

Our file formats will exist both in non-proprietary and proprietary formats. The non- proprietary formats will ensure that these data are able to be used by anyone wishing to do so once they are deposited and made openly available.

Surveys will exist in .csv (non-proprietary), MS Excel, & SPSS (both proprietary) formats. For more information regarding SPSS see: SPSS Wikipedia https://en.wikipedia.org/wiki/SPSS

Interviews & focus groups data will exist in .mp3 (non-proprietary), MS Word & NVivo (both proprietary) formats. For more information regarding NVivo see: NVivo Wikipedia https://en.wikipedia.org/wiki/NVivo

Any survey data deposited for sharing and long-term access will be in .csv format so that anyone can use them without requiring proprietary software.

The final de-identified versions of the interviews and focus groups transcripts will be exported into a basic non-proprietary text format for deposit, long-term preservation and access.

If data are collected using laptops or mobile devices, please explain how you will securely store and transfer the data.

Laptops are not being used for any data collection, though encrypted digital voice recorders (DVRs) will be used to collect both interviews and transcripts. Interviews and focus group digital audio files will not be stored on the DVRs, only collected and then securely transferred to the project's cloud based virtual research environment space via a secure FTP (File Transfer Protocol).

2

MODÈLES DE PGD

Type de recherche	Titre
Sciences humaines numériq	Plan de gestion des données pour le projet Belges et Français dans les Prairies
	Étude du paysage sonore du silence (Modèle)
	Sciences Humaines Numériques
	Sciences humaines numériques et données secondaires
Méthodes mixtes	Modèle fictif sur « Méthodes mixtes »

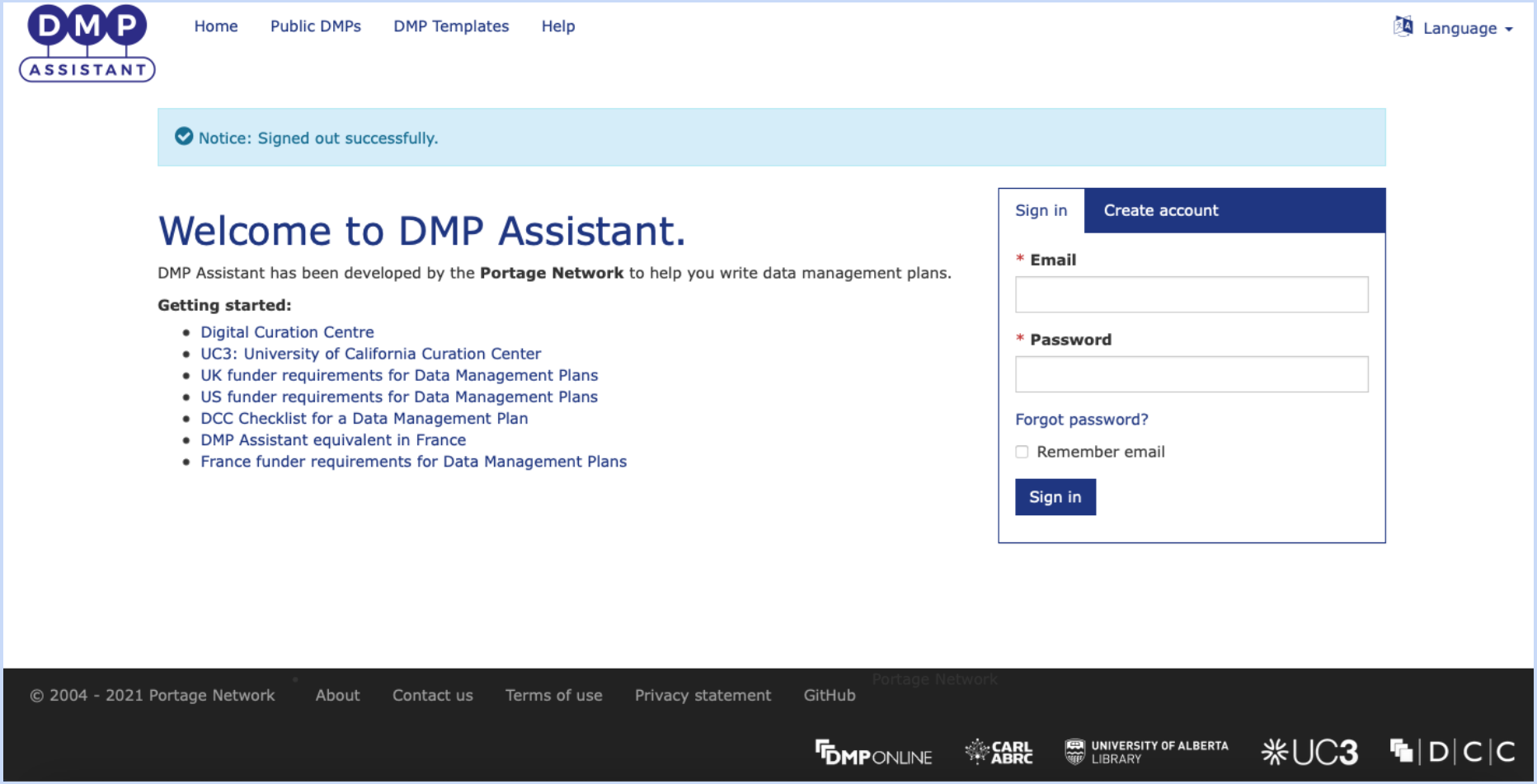
MODÈLES DE PGD

Type de recherche	Titre
Sciences naturelles	Plan de gestion des données du groupe de recherche en écohydrologie (Modèle)
	Plan de gestion des données pour Reproductibilité computationnelle dans l'informatique haute performance (Modèle)
Sciences sociales	Plan de gestion de données pour Personnes, places, politiques et perspectives (Modèle)
	Plan de gestion des données pour Utilisation des sites web de profils universitaires (Modèle)

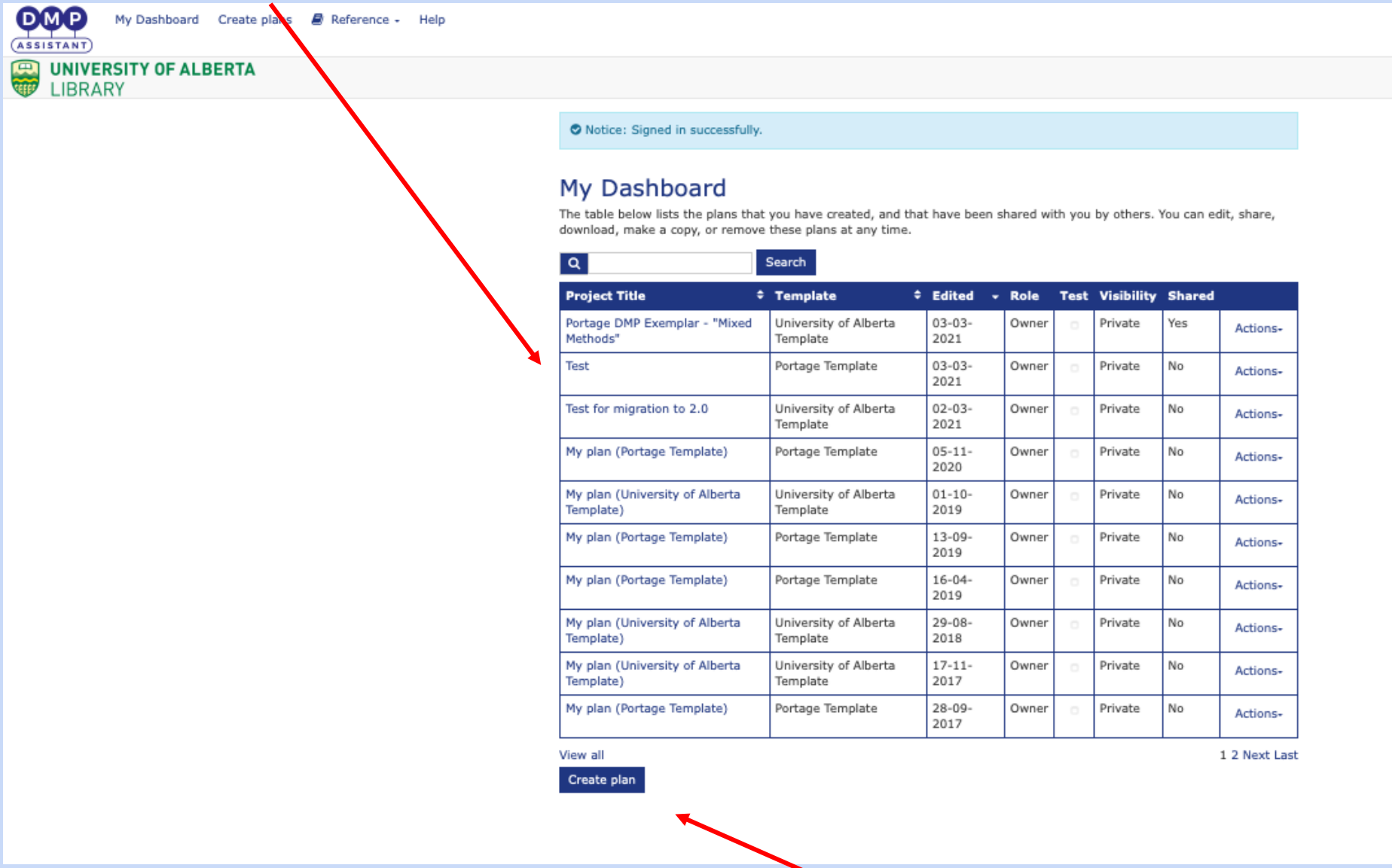
DMP Assistant Key Features

- Supports cross institutional research collaboration
- Freely available to end users
- Web-based, bilingual tool
- Endorsed by the Tri-Agencies
- Guidance & Examples
- Can have multiple DMPs
- Shared access - co-owner, editor, or read-only
- Exportable (e.g., pdf, text, MS Word)
- Living document - modifiable when needed

DMP Assistant

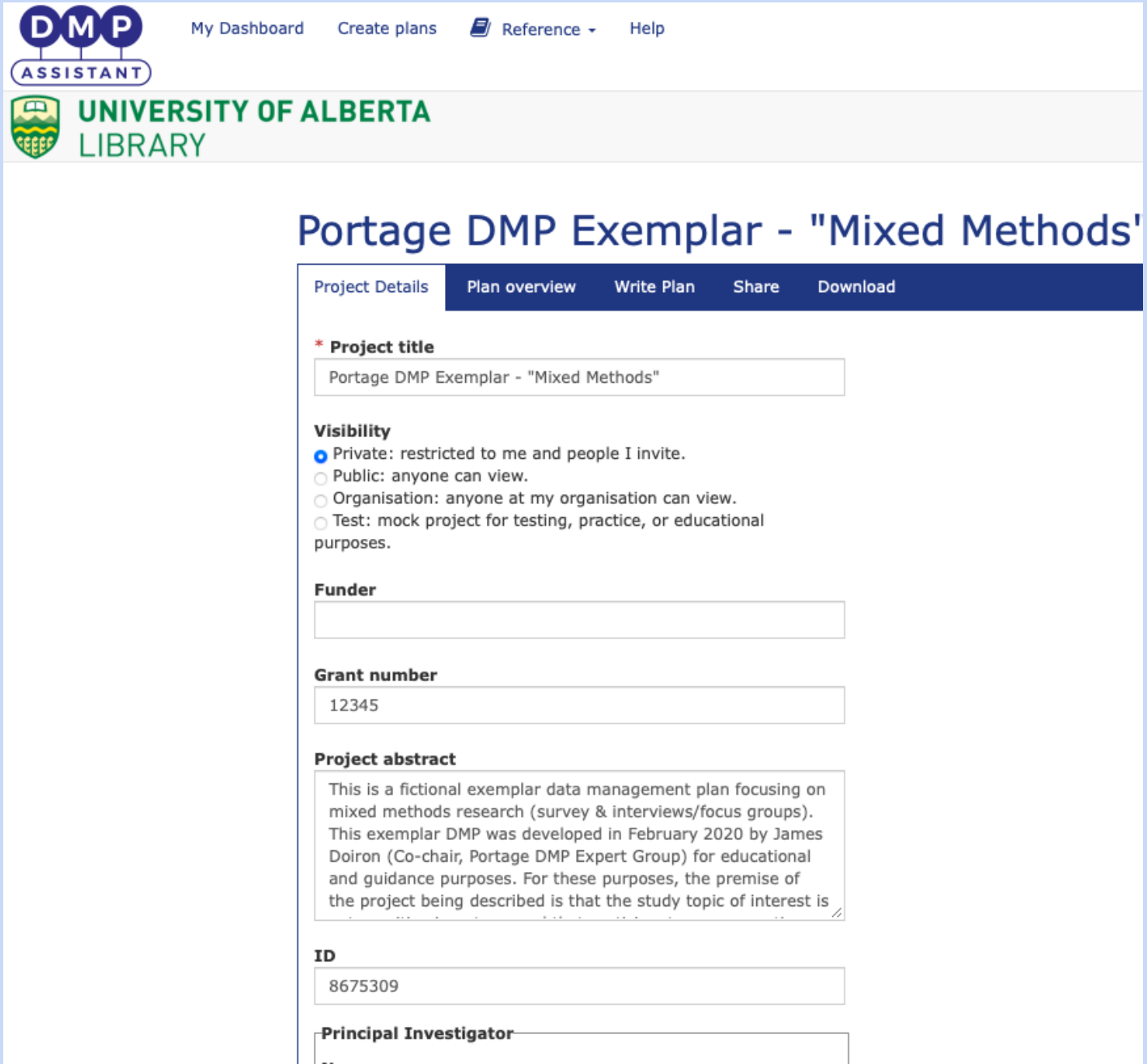


Upon signing in, you can either access an existing DMP…..

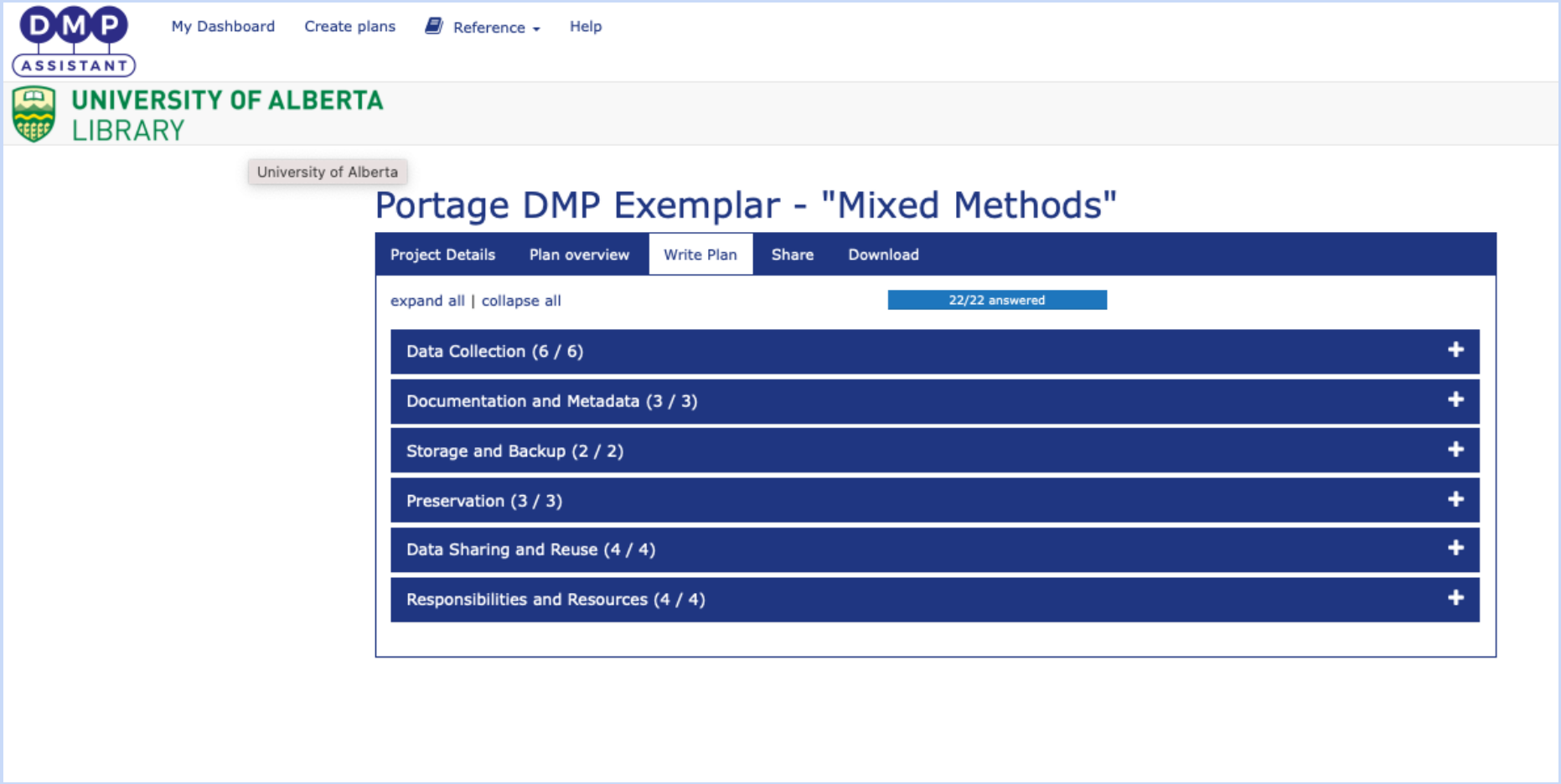


…..or create a new one

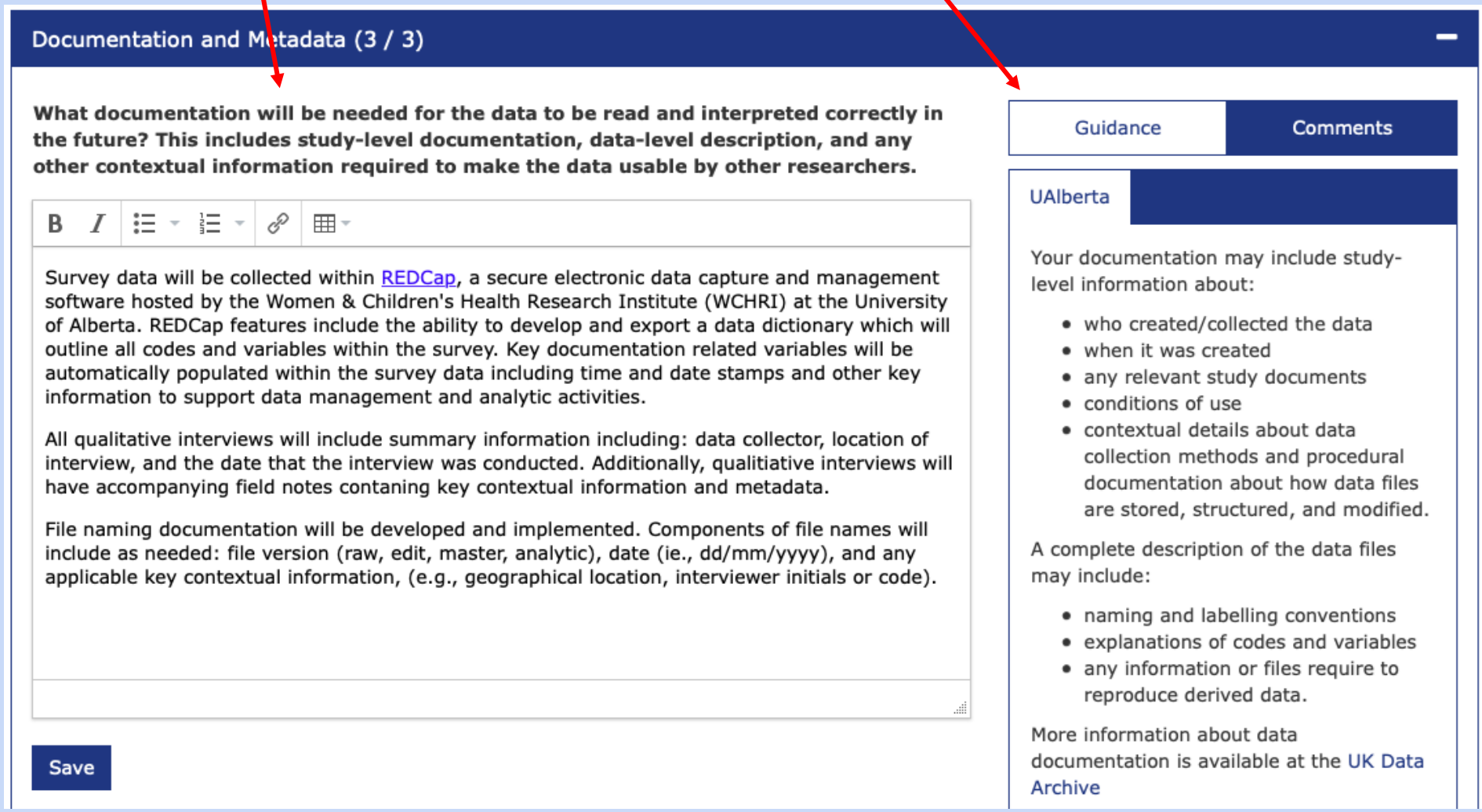
Each DMP has study level information associated with it



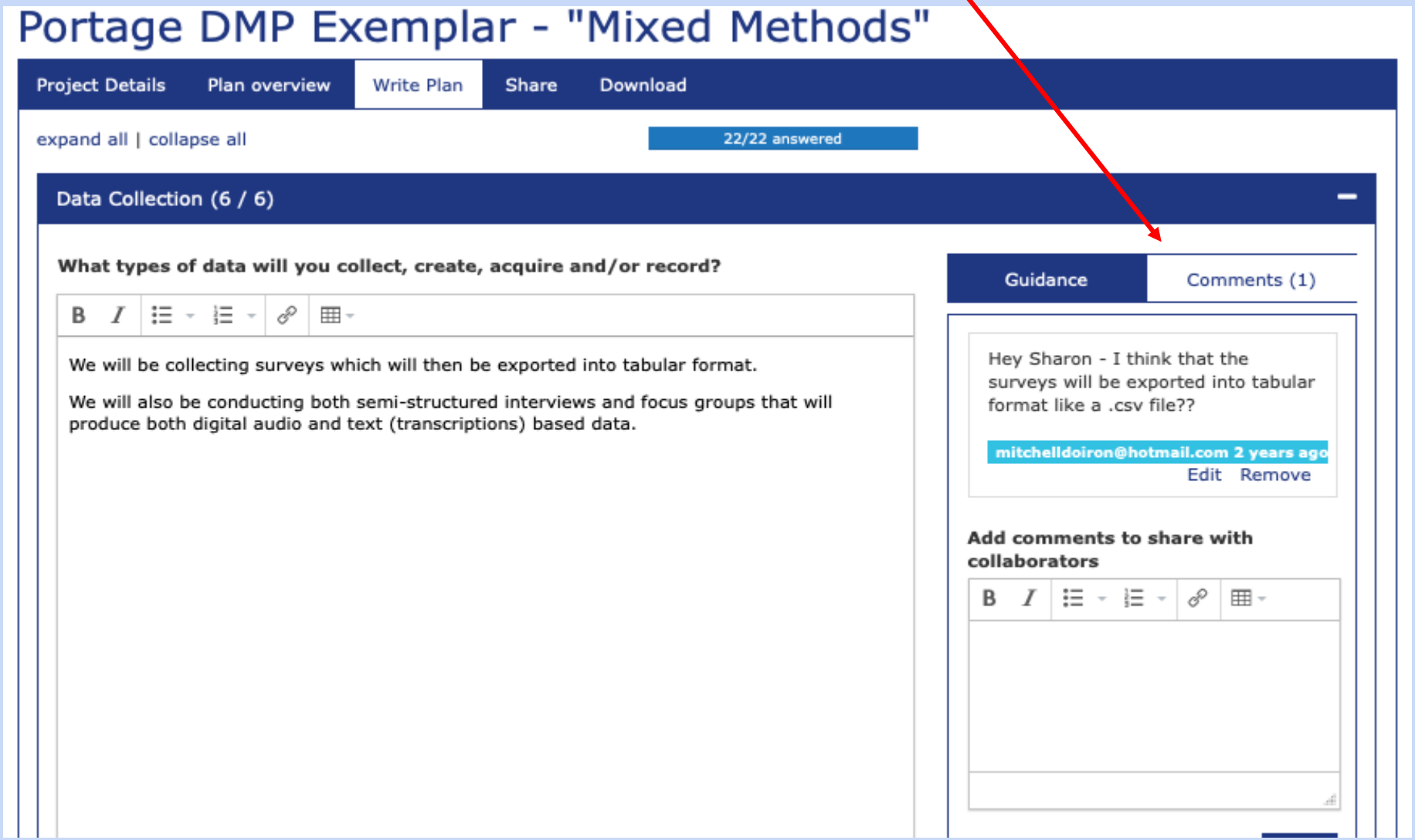
You can choose to answer questions within any given section at any time...



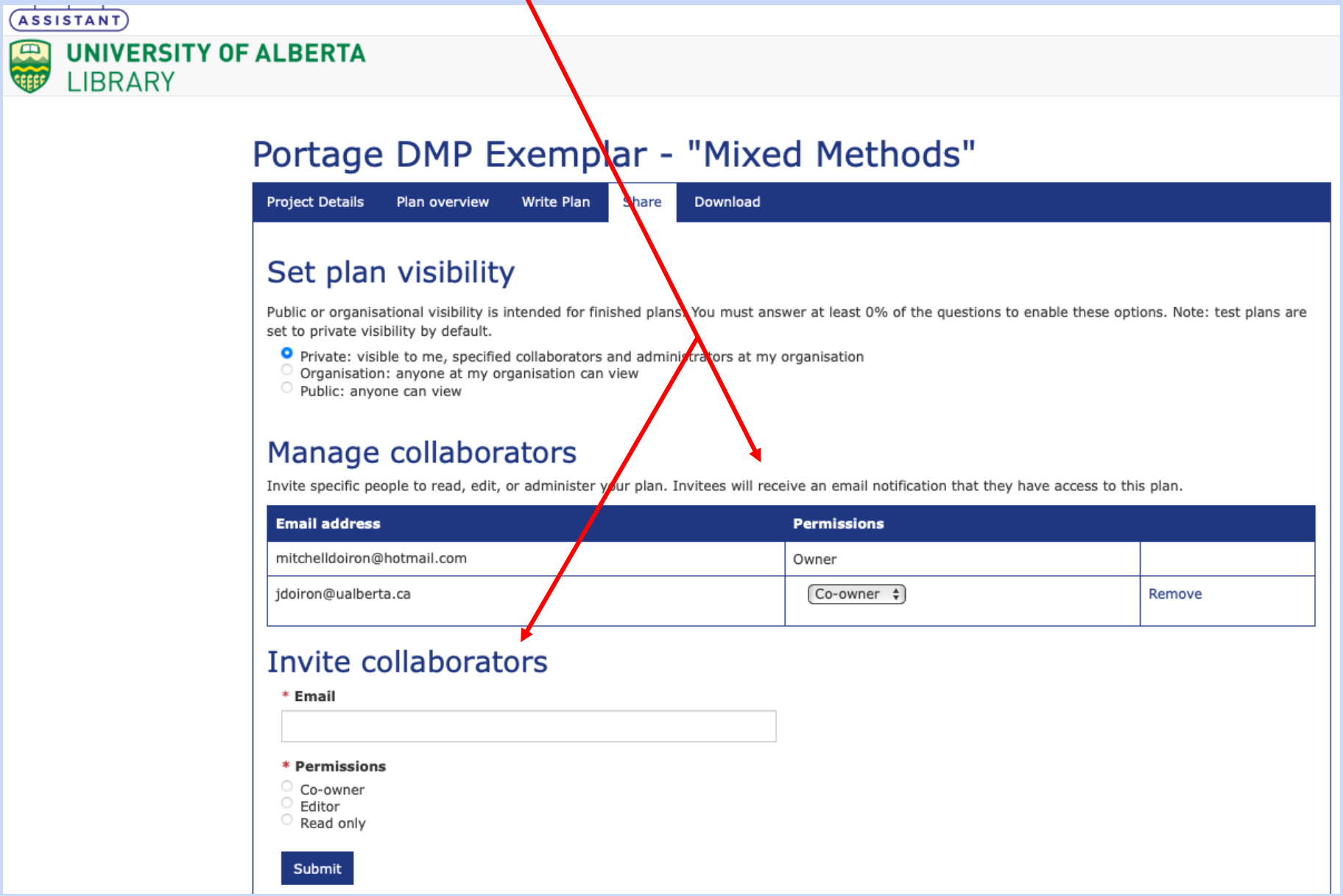
Dedicated space as well as guidance customized guidance is provided for each question...



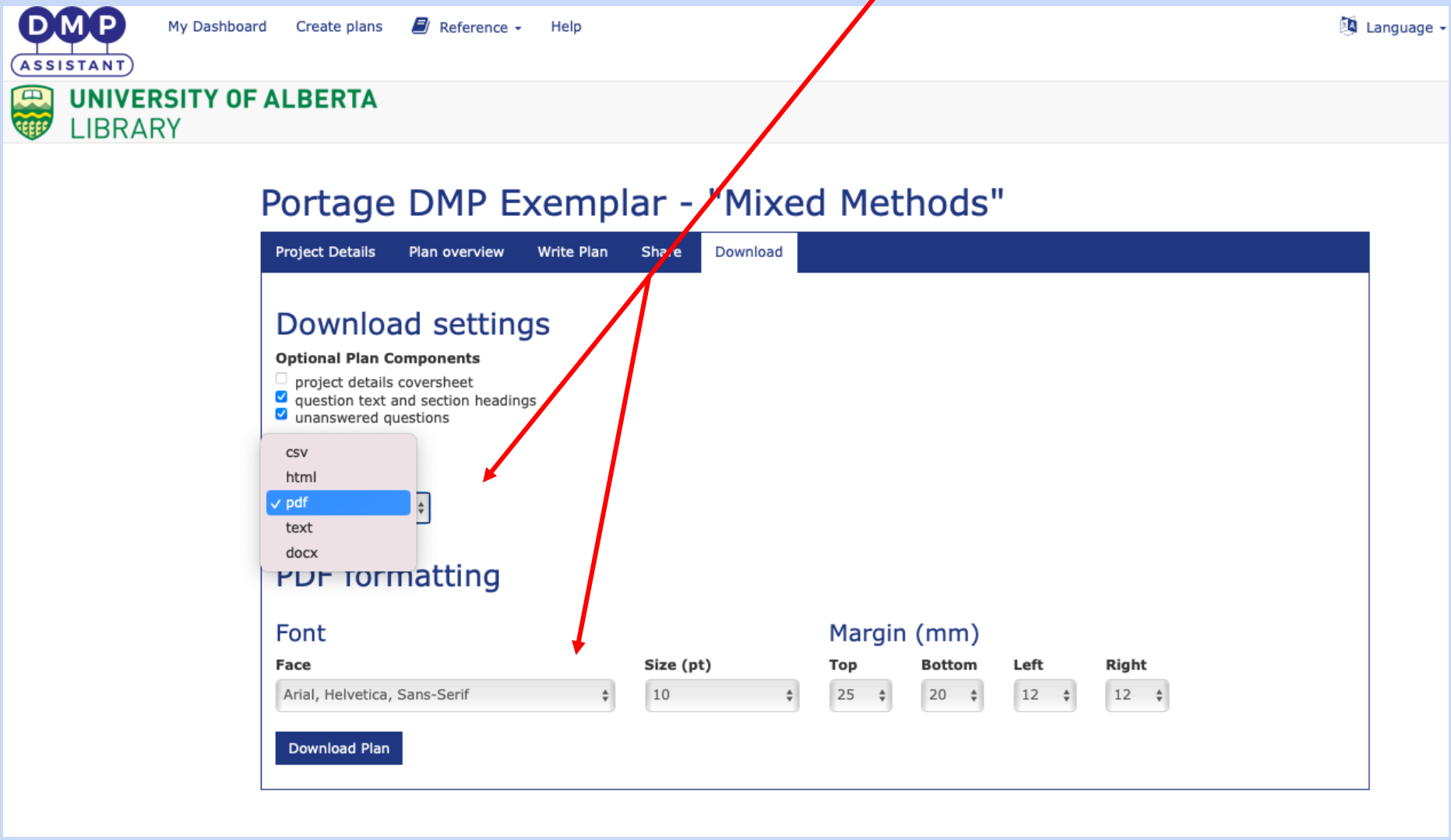
For any given question you are able to have ‘share notes’ with your collaborators



You can easily add collaborators to your DMP, with different permissions settings...



A range of options are available both for formatting and downloading your DMP...



Moving forward….

- **DMP Assistant**
  - ➔DMP Assistant 2.0 Bilingual webinars soon to come - see Portage events page
  - ➔New discipline and methodology specific templates
  - ➔More DMP exemplars to come and based on the new templates
  - ➔More features and support for machine actionable DMPs

- **DMP Assistant Steering Committee**
  - ➔consists of a wide range of stakeholders (e.g., Portage, host institution, researchers, Tri-Agencies, First Nations Information Governance Centre)
  - ➔Helps to guide the ongoing development and sustainability of DMP Assistant
  - ➔Helps to define and prioritize work flow and roadmap activities

- **DMP Expert Group**
  - ➔New DMPEG members
  - ➔DMP Assistant 2.0 tutorial video working group
  - ➔DMP Assistant 2.0 customization guide working group

Portage Webinar – Support Your Research with DMP Assistant 2.0!

Date: March 30, 2021

Time: 1 pm ET

Location: Online

Registration

Data management plans, or DMPs, are one of the foundations of good research data management. Hosted by the University of Alberta Library and supported by the Portage Network, the DMP Assistant is a national, open, bilingual data management planning (DMP) tool to help researchers better manage their data throughout the lifespan of a project. The tool develops a DMP by prompting researchers to answer a number of key data management questions, supported by best-practice guidance and examples. Building on the preceding DMP-focused webinar, this session will be of interest to researchers, graduate students, librarians, and research support stakeholders. Participants will take an in-depth look at the newly launched DMP Assistant 2.0, including all of its enhanced key features for both end-users and institutional administrators, as well as a brief look at the future of the platform.

Speaker:

Robyn Nicholson, Data Management Planning Coordinator, Portage Network

This webinar will be presented in English with simultaneous interpretation in French and will be recorded. Engagement during the webinar in both official languages is welcomed. Previous webinar recordings are available on the CARL YouTube Channel. Links to the recordings and slides can also be found on Portage Training Resources.

Register here!

UNIVERSITY OF ALBERTA
LIBRARIES

DMP ASSISTANT

portage
SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE
SHARED STEWARDSHIP OF RESEARCH DATA

Questions & Discussion

James Doiron
RDM Services Coordinator, UofA Library
Academic Director, UofA Research Data Centre
Co-Chair, Portage DMP Expert Group

jdoiron@ualberta.ca