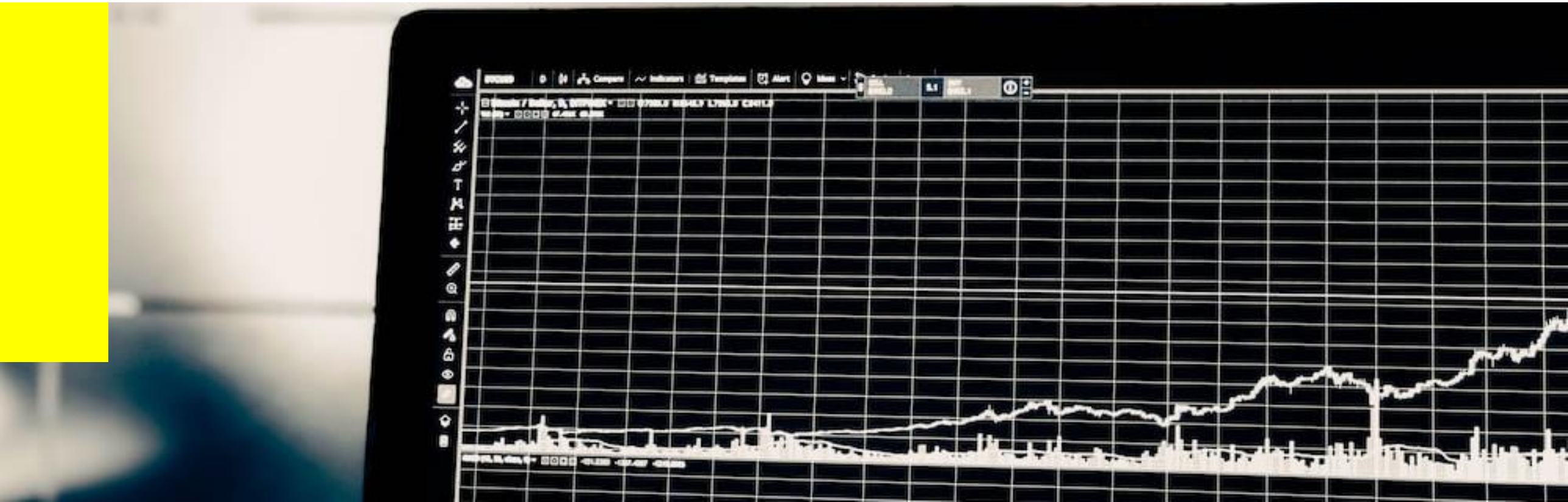
University of Applied Sciences and Arts Northwestern Switzerland
School of Life Sciences

S4 – Data Stewardship I

D4 – Data and Ethics



Schedule

KW		Date	#	Topics	LernSetting WI	Lecturer
38 39	Self Study	First 2 weeks	0	Awareness - Entry Test with Moodle Test (20% counted to course grade)	Virtual	Selfstudy
38		KW38	0 + 7	Coaching Session (according to the information of the respective school)	on site	JRN= Juchler Norman Rerabek Martin Nyfeler Matthias
38	Fr, afternoon	23.09.2022	1	Personal Security	Virtual	Pascal Moriggl
39		KW39	1	Coaching Session	on site	FHNW: Pascal Moriggl ZHAW: JRN
39	Fr, afternoon	30.09.2022	2	Information Security & Cybersecurity I	Virtual	Petra M. Asprion
40		KW40	2	Coaching Session	on site	FHNW: Petra M. Asprion ZHAW: JRN
40	Fr, afternoon	07.10.2022	3	Information Security & Cybersecurity II	Virtual	Petra M. Asprion
41		KW41	3	Coaching Session	on site	FHNW: Pascal Moriggl ZHAW: JRN
41	Fr, afternoon	14.10.2022	4	Data Stewardship I	Virtual	Pascal Moriggl
42		KW42	4	Coaching Session	on site	FHNW: Pascal Moriggl ZHAW: JRN
42	Fr, afternoon	21.10.2022	5	Data Stewardship II	Virtual	Pascal Moriggl
43		KW43	5	Coaching Session	on site	FHNW: Pascal Moriggl ZHAW: JRN
43	Fr, afternoon	28.10.2022	6	Data Ethics	Virtual	Pascal Moriggl
44		KW44	6	Coaching Session	on site	FHNW: Pascal Moriggl ZHAW: JRN
44	Fr, afternoon	04.11.2022	7	Data Privacy	Virtual (Flipped Classroom)	Pascal Moriggl

Agenda

Part I : Repetition Last Week

Part I : Introduction Data Stewardship

Part II : Data Management Plan

Part III : DMP Tool

Lecturer 3
GRC

GRC– Why of relevance?

Our topic --
Governance, Risk and Compliance – GRC (I&CS)

Lecturer 3 GRC

Frameworks Standards Guidelines

Translating it to the organization

1 Regulations

2 Implementation of Frameworks

3 Policy & Procedure Implementation

Evidence Collection

SOX

HIPAA

PCI DSS

..

COSO ERM

COBIT

ISO 20000-x

ISO 27000x

CMMI

ISO 31000

Business Risks

Business Controls:
- Change Controls
- Risk mitigation
- Access controls and SoD
- Documentation
- …

Establish | Document | Test

Demonstration of compliance – (Achieve Evidence)

People | Processes | Technology | Facilities | Data

→ ES/ERP relevant

→ ES/ERP relevant

→ ES/ERP relevant

Lecturer 3
GRC

CLASSIFICATION				
L1 Information intended and released for public use.	**L2** Information that may be shared only within the Harvard community.	**L3** Confidential and sensitive information, intended only for those with a "business need to know."	**L4** High-risk information that requires strict controls.	**L5** Extremely sensitive information requiring specific controls and isolation from the network.
The University intentionally provides this information to the public.	The University chooses to keep this information private, but its disclosure would not cause material harm.	Disclosure of this information beyond intended recipients might cause material harm to individuals or the University.	Disclosure of this information beyond specified recipients would likely cause serious harm to individuals or the University.	Disclosure of this information could cause criminal liability; loss of insurability or employability; or severe social, psychological, reputational, financial, or other harm to an individual or group.
Examples • Published research • Course catalogs • Published faculty and staff information • Student directory information* • Basic emergency response plans (life safety) • University-wide policies • Harvard publications • Press releases • Published marketing materials • Regulatory and legal filings • Published annual reports • Code contributed to Open Source • Released patents • Plans of public spaces *Directory information about students who have requested FERPA blocks must be classified and handled as L3, at minimum.	**Examples** • Department policies and procedures • Employee web/intranet portals • Harvard training materials • Pre-release articles • Drafts of research papers • Work papers • Patent applications • Grant applications • Non-public building plans or layouts (excluding L3 or L4 items) • Information about physical plant (excluding L3 or L4 items)	**Examples** • Non-directory student information • Non-published faculty and staff information • Information protected under FERPA, in general • HUID tied to an individual • Personnel records** • Donor information (excluding L4 data points or special handling) • Non-public legal work and litigation information • Budget /financial transactions information • Non-public financial statements • Information specified as confidential by vendor contracts and NDAs • Information specified as confidential by Data Use Agreements • General security findings or reports (e.g. SSAE16) • Most Harvard source code • Non-security technical specifications/architecture schema • Library/museum object valuations • IRB records	**Examples** • Passwords and PINs • System credentials • Private encryption keys • Government issued identifiers (e.g. Social Security Number, Passport number, driver's license) • Individually identifiable financial account information (e.g. bank account, credit or debit card numbers) • Individually identifiable health or medical information*** • Individually identifiable research data • Details of significant security exposures at Harvard (e.g. vulnerability assessment and penetration test results) • Security system procedures and architectures • Trade secrets • Systems managing critical Operational Technology	**Examples** • Research data classified as Level 5 by the IRB • Information or research under a contract stipulating specific security controls beyond L4

Introduction

Learning Goals

- ✓ Increase awareness of data trends and the related need for order
- ✓ Understand the difference and link between data governance and data stewardship
- ✓ Know the key principles of data stewardship

Relevance

Open Data

The benefits and value of open data

Building on the Digital Single Market efforts: open data in Europe today

22/01/2020 Europe

Open data in Europe

Most national governments in Europe have developed an open data agenda and have established open data portals supported by solid open data policies and strategies. In 2019, an overall maturity of 74% was achieved by the EU28 in terms of policy maturity (visible on the ODM dashboard), indicating that Member States have developed a strong foundation in terms of their open data policy framework. The increased amount of national data portals is reflected in the evolution of datasets available on the European Data Portal (EDP). For example, in May 2016 the EDP only had over 400,000 available datasets, which grew to over 890,000 in August 2019 and is expected to grow further.

https://data.europa.eu/en/datastories/benefits-and-value-open-data

Relevance

Open Data = Economic Growth, Positive Impact on Society

Increasing the quality, efficiency, and transparency of public services;

Cost saving, as it is for example forecasted that national governments of the EU28+ could save 1,7 billion euro by 2020, Link opens in a new window; and

Greater efficiency in processes and delivery of public services. This can be illustrated by an example from the Netherlands, Link opens in a new window, where the Ministry of Education publishes education-related data for re-use. Since then, the number of questions they receive has dropped, reducing workload and costs. The remaining questions are now also easier for civil servants to answer because it is clear where the relevant data can be found.

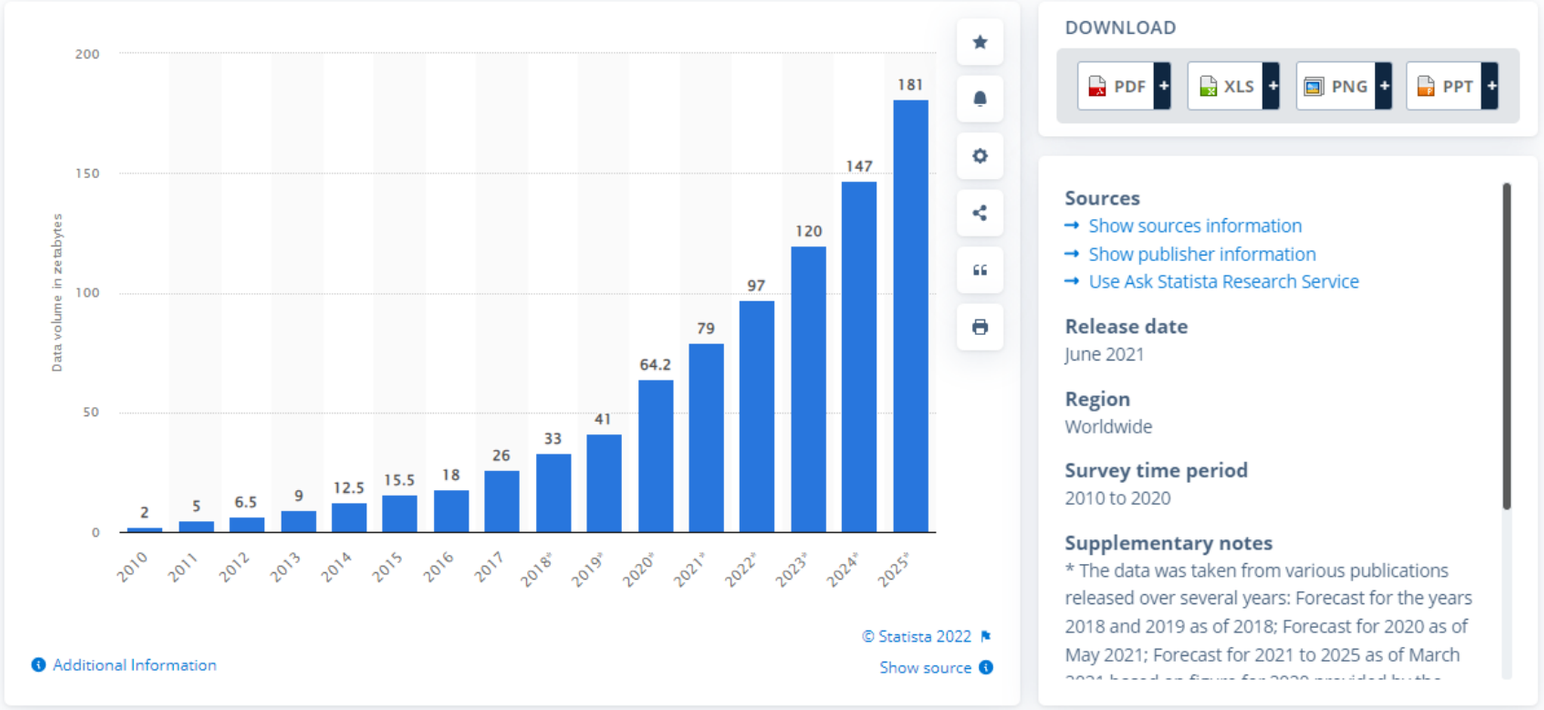
It is estimated that by 2020, Link opens in a new window, the market size for open data has increased by 36.9%, to a value of 75.7 billion EUR. When looking at the impact of open data in a specific sector, the public sector is expected to have the highest share in terms of direct market size with a value of 22.11 million EUR, Link opens in a new window. The growth of the open data market size is also expected to trigger a higher demand for skilled open data workers. It is forecasted that open data has created 100,000 jobs by 2020.

https://data.europa.eu/en/datastories/benefits-and-value-open-data

Relevance

Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025

(in zettabytes)



DOWNLOAD

PDF XLS PNG PPT

Sources
- Show sources information
- Show publisher information
- Use Ask Statista Research Service

Release date
June 2021

Region
Worldwide

Survey time period
2010 to 2020

Supplementary notes
* The data was taken from various publications released over several years: Forecast for the years 2018 and 2019 as of 2018; Forecast for 2020 as of May 2021; Forecast for 2021 to 2025 as of March

Zettabyte Defined
A zettabyte is a measure of digital storage capacity. A zettabyte is read as the 2 to the 70th power bytes. It is also equal to a thousand exabytes, a billion terabytes or a trillion gigabytes. Simply, it would mean one billion, one terabyte hard drives would be needed to store one zettabyte of data.

Due to the zettabyte unit of measurement being so large, it is only used to measure large aggregate amounts of data. Even all the data in the world is estimated to be only a few zettabytes.

A byte is a data measurement unit that contains eight bits, or a series of eight zeros and ones. There is nine types of bytes.

Byte
Kilobyte
Megabyte
Gigabyte
Terrabyte
Petabyte
Exabyte
Zettabyte
Yottabyte

https://www.indicative.com/resource/zettabyte/

https://www.statista.com/statistics/871513/worldwide-data-created/

Relevance

Review > Drug Discov Today. 2020 Sep;25(9):1624-1638. doi: 10.1016/j.drudis.2020.07.005. Epub 2020 Jul 11.

Advancing computer-aided drug discovery (CADD) by big data and data-driven machine learning modeling

Linlin Zhao [1], Heather L Ciallella [1], Lauren M Aleksunes [2], Hao Zhu [3]

Affiliations + expand

PMID: 32663517 PMCID: PMC7572559 DOI: 10.1016/j.drudis.2020.07.005

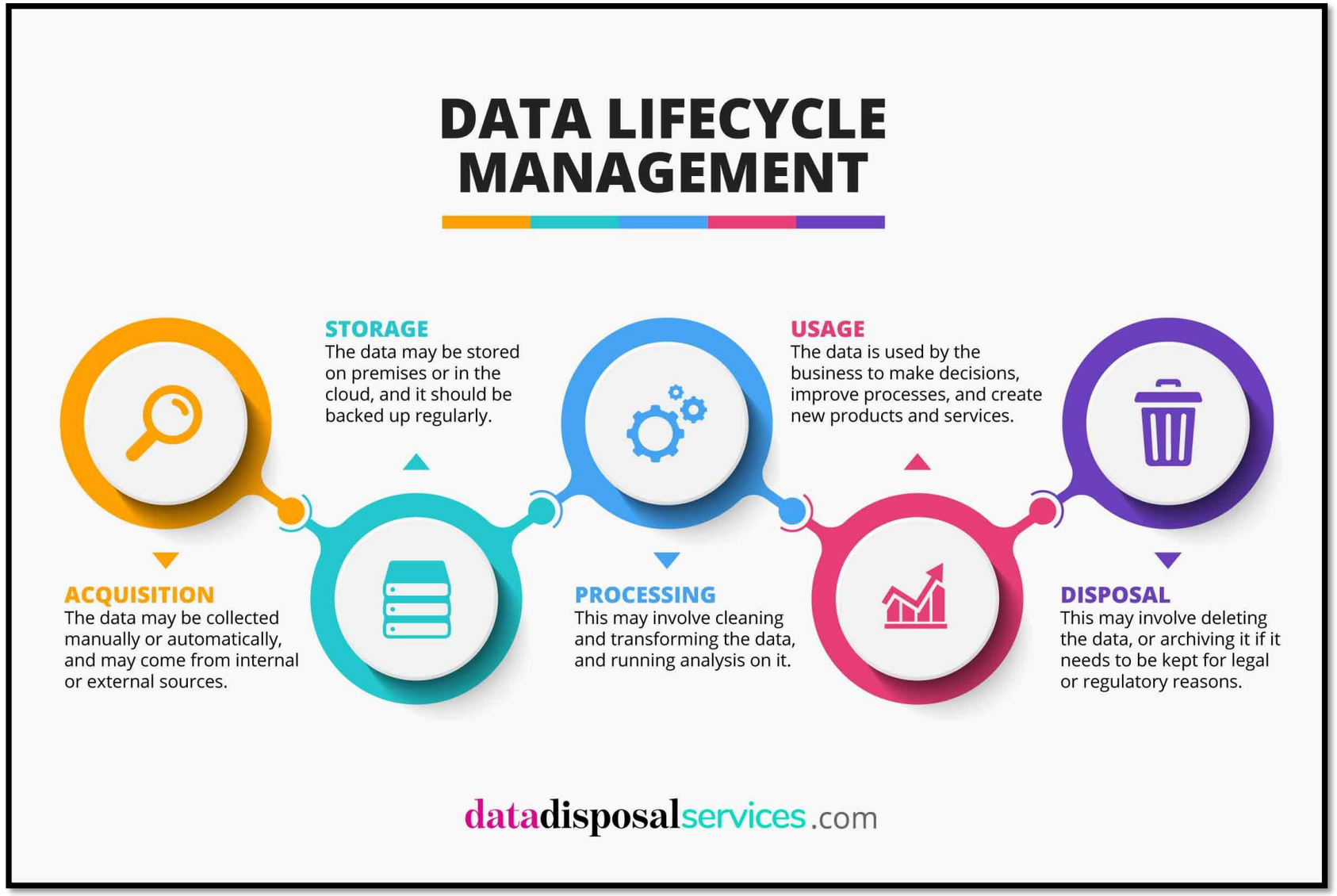
Free PMC article

Abstract

Advancing a new drug to market requires substantial investments in time as well as financial resources. Crucial bioactivities for drug candidates, including their efficacy, pharmacokinetics (PK), and adverse effects, need to be investigated during drug development. With advancements in chemical synthesis and biological screening technologies over the past decade, a large amount of biological data points for millions of small molecules have been generated and are stored in various databases. These accumulated data, combined with new machine learning (ML) approaches, such as deep learning, have shown great potential to provide insights into relevant chemical structures to predict in vitro, in vivo, and clinical outcomes, thereby advancing drug discovery and development in the big data era.

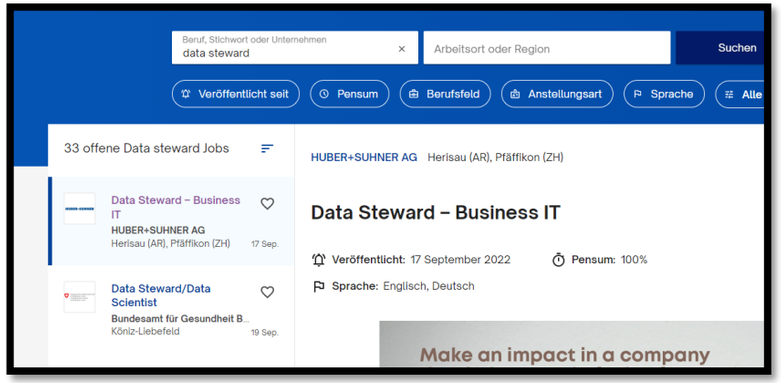
https://pubmed.ncbi.nlm.nih.gov/32663517/

Relevance


DATA LIFECYCLE MANAGEMENT
ACQUISITION
The data may be collected manually or automatically, and may come from internal or external sources.
STORAGE
The data may be stored on premises or in the cloud, and it should be backed up regularly.
PROCESSING
This may involve cleaning and transforming the data, and running analysis on it.
USAGE
The data is used by the business to make decisions, improve processes, and create new products and services.
DISPOSAL
This may involve deleting the data, or archiving it if it needs to be kept for legal or regulatory reasons.
datadisposalservices.com

Relevance

„Dear XY, You will become our new, internal data steward! Good Luck!“



Data Steward

Competencies

Departmental knowledge

Deep knowledge of the operational area they are responsible for, including understanding processes, rules and requirements. As well as functional understanding, a knowledge of the data flows and data sources will be important.

Communication skills

The ability to interpret and communicate policy or business rules to end users. At the same time being able to work with the technology and policy owners to ensure ideas are fed back.

Collaboration skills

Working with other Data Stewards and stakeholders across the organization to ensure data flows smoothly around the organization.

https://www.data-vault.co.uk/qualities-data-steward/

Data Steward

Responsibilities

- Ensuring data quality, data definition and privacy standards are met.
- Ensuring that data is fit for purpose (including completeness, accuracy and integrity).
- Managing metadata and processes to ensure proper use of data being read, created, collected, reported, updated or deleted.
- Ensuring data is protected and security procedures are enforced.
- Take an active part in the data governance framework to feedback on existing practise and recommend improvements.

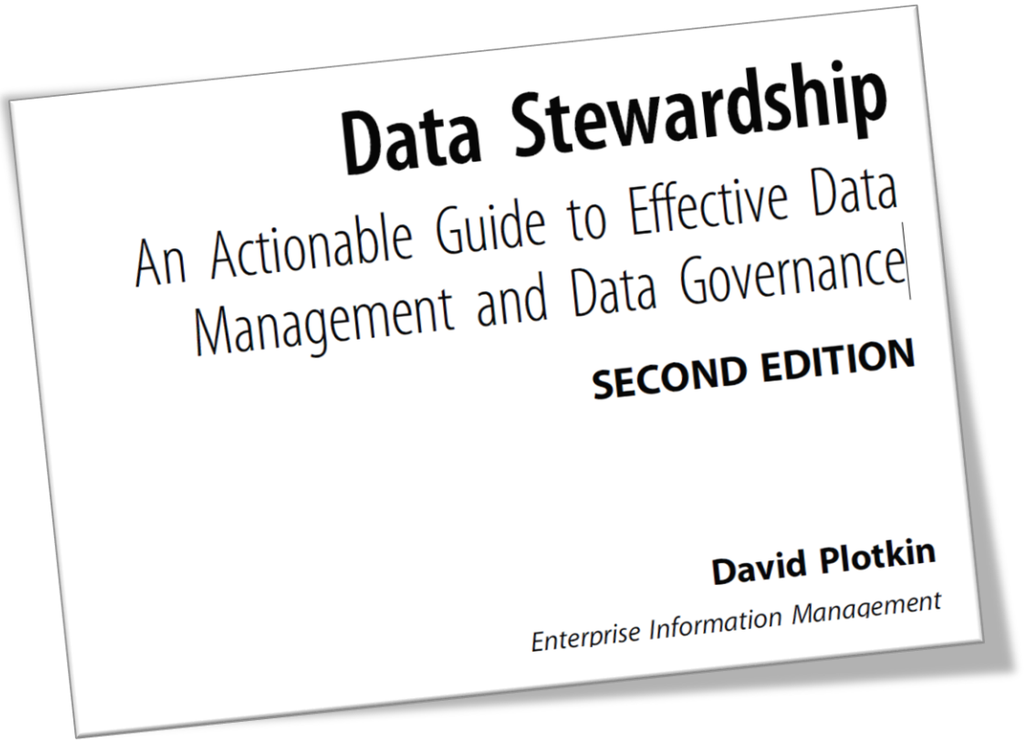
https://www.data-vault.co.uk/qualities-data-steward/

Data Stewardship

Data Stewardship

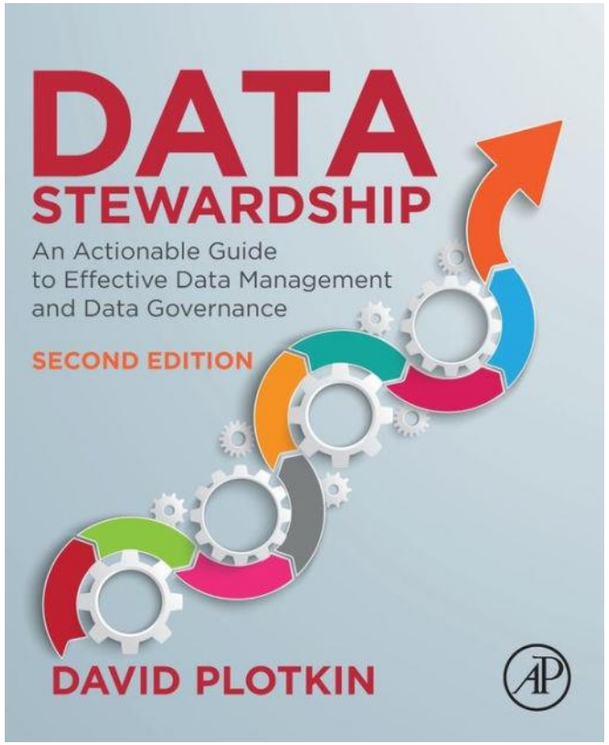
"Data stewardship is concerned with the science and practice of data collection for the purposes of analysis, reflecting the values of fair information practices. In practice data stewardship **is a collection of methods and mechanisms of data management** encompassing acquisition, storage, protection, aggregation, deidentification, and procedures for data release, use, and re-use, to ensure that the data assets are of high quality, easily accessible, and used appropriately."

https://unece.org/



Data Governance

"Data governance concerns decision-making and authority for data related matters, whether within or between enterprises and public agencies. According to a seminal text on data stewardship, data governance is **a system of decision rights and accountabilities for information related processes**, executed according to agreed-upon models governing the kind of data stored, the authority to access data, and the methods of data access."



Plotkin, D. (2021). Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance (2nd Ed.). London, UK: Academic Press.

Data Stewardship vs. Data Governance?

Data Governance is about **how** people manage and make decisions about data rather than about the data itself, whereas

Data Stewardship is about ensuring that people are properly organized and do the right things **to make their data understood**, trusted, of high quality, and ultimately, suitable and usable for the enterprise's purposes.

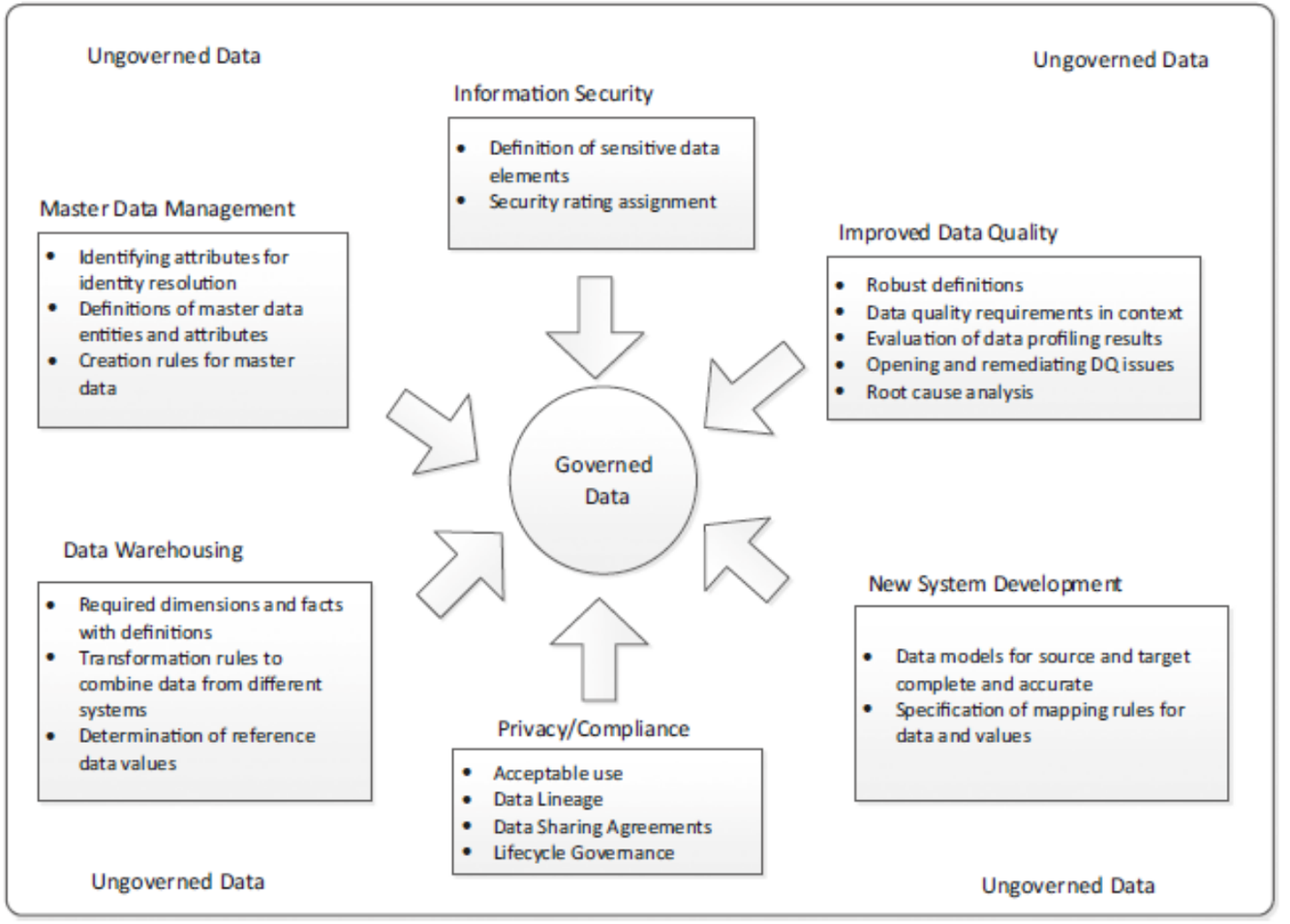
https://www.sciencedirect.com/science/article/pii/B9780128221327000012?via%3Dihub

Context

- exponential growth in data generated worldwide (e.g. IoT) and the arrival of new technologies (AI, Edge Computing) increases importance on data
- use of electronic devices through communication networks has led to rich data pools discover patterns, infer indicators, develop business models, ….
- digitalization of society and economy has placed data access and sharing at the core of innovation and public trust
- there are huge possibilities for new types of data services, more timely and granular data, new insights by linking data from different sources and topics
- but there are also huge risks: data could be used unethically, the '***digital divide***' could become an 'information divide', invasion of privacy

https://unece.org/

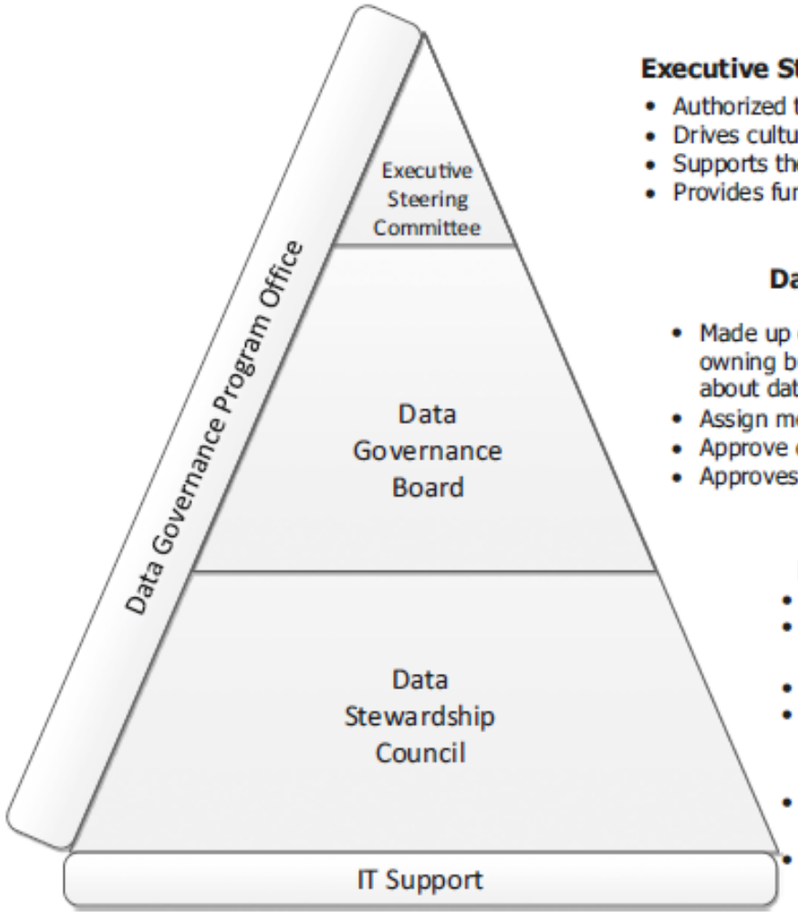
Data Stewardship: Drivers



There are many drivers for moving data from an ungoverned state to a governed state.

Plotkin, D. (2021). Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance (2nd Ed.). London, UK: Academic Press.

Data Stewardship: Organisation



Executive Steering Committee

- Authorized to change the organization
- Drives cultural change
- Supports the program enterprise-wide
- Provides funding for the Data Governance Program

Data Governors

- Made up of high-ranking representatives of data-owning business functions who can make decisions about data for the company.
- Assign members of Data Stewardship Council
- Approve decisions of Data Stewardship Council
- Approves data-related policies

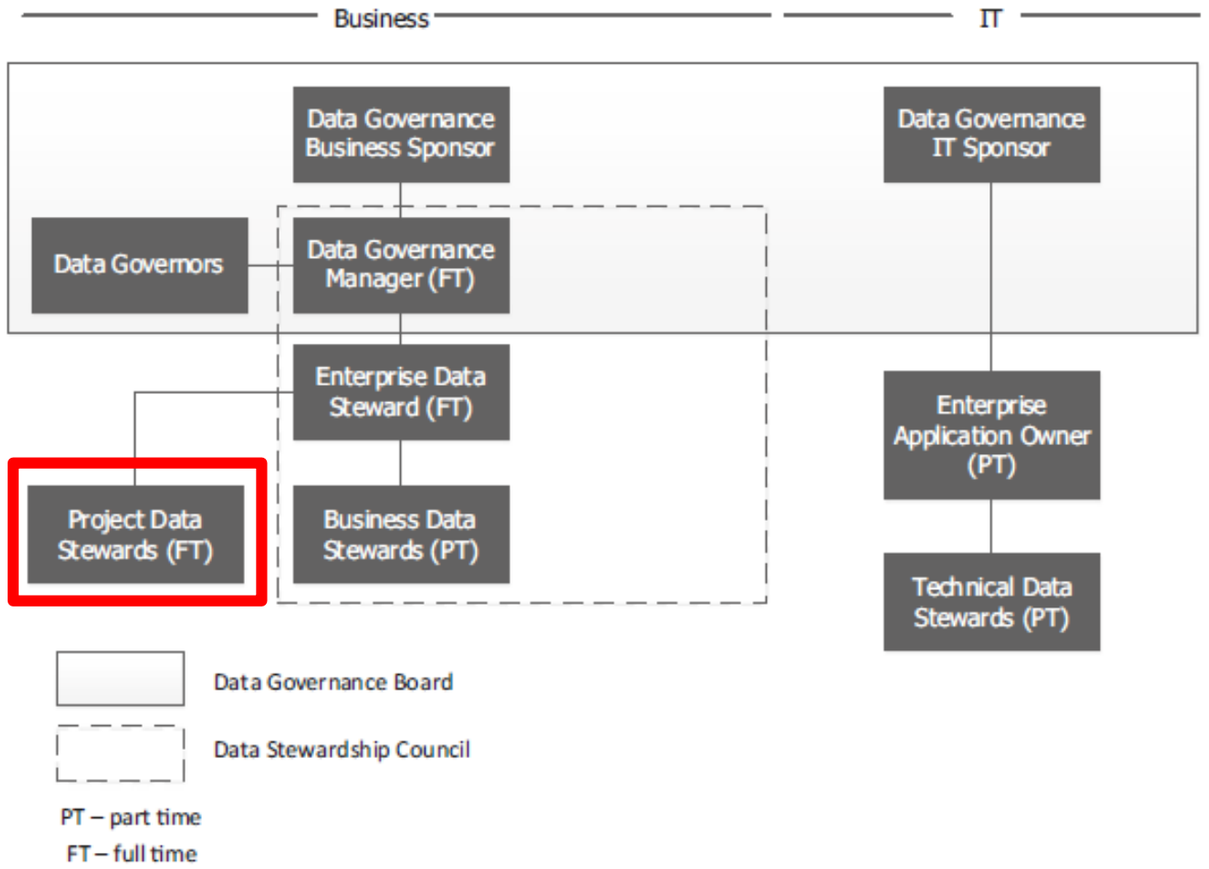
Business Data Stewards

- Experts on use of their data.
- Able to reach out to SMEs to gather information and make decisions
- Knows who their stakeholders are
- Typically someone who others come to as the most knowledgeable about the meaning of the data (and how it is calculated).
- Makes recommendations on data decisions and write data-related procedures.
- May be part of data domain steward team

A Data Governance Program is often arranged as a pyramid, with support from IT and a Data Governance Program Office.

Plotkin, D. (2021). Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance (2nd Ed.). London, UK: Academic Press.

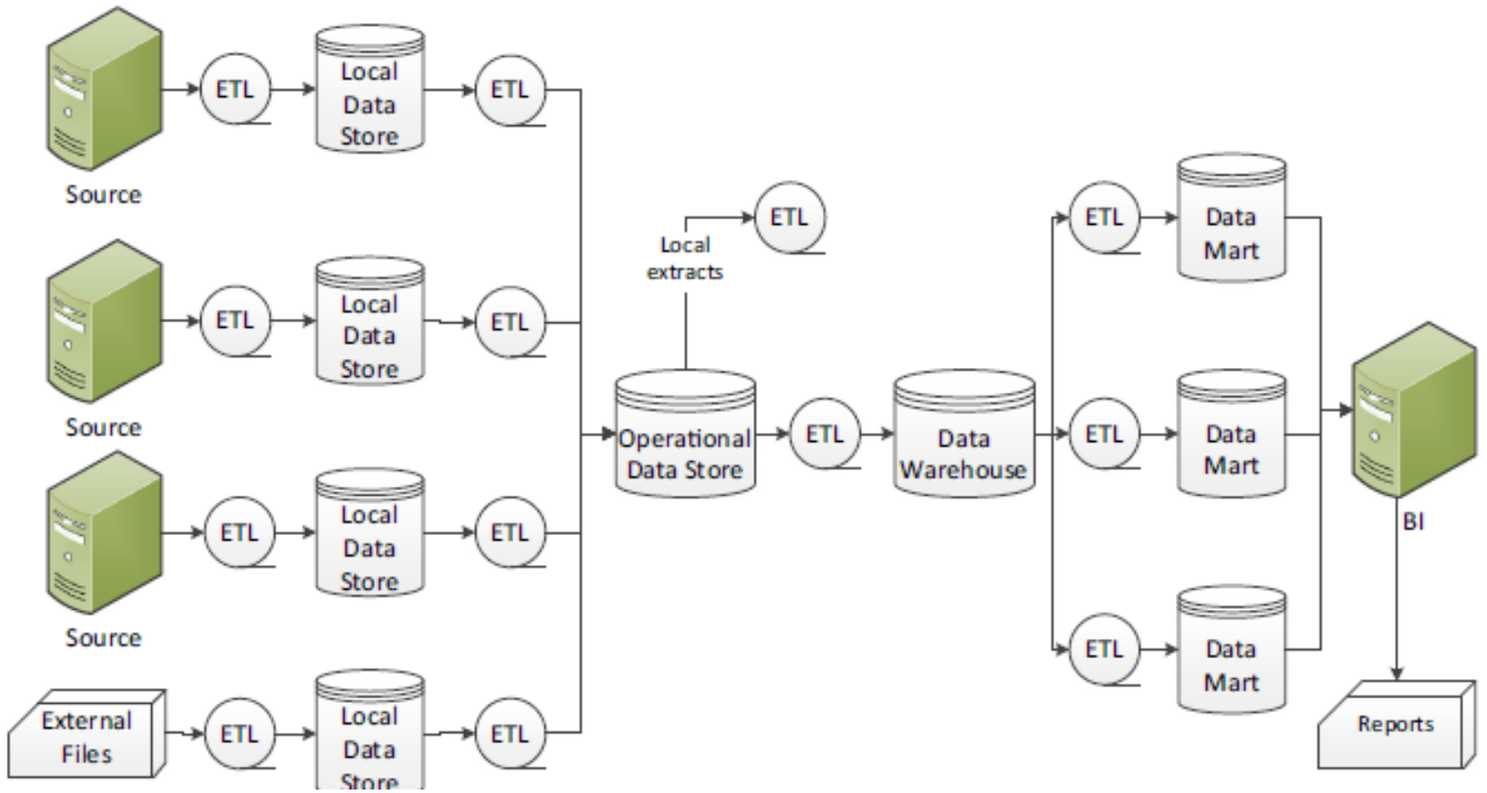
Data Stewardship: Business / IT Alignment



The Business and IT view of the Data Governance Organization.

Plotkin, D. (2021). Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance (2nd Ed.). London, UK: Academic Press.

Data Stewardship: The „Bigger“ Flow-Picture



Information flows through an enterprise in the “Information Chain.”

Plotkin, D. (2021). Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance (2nd Ed.). London, UK: Academic Press.

Let‘s get straight to the

Chapter 6

Practical Data Stewardship

Rancourt, E. (2019). The scientific approach as a transparency enabler throughout the data life-cycle. *Statistical Journal of the IAOS,* 35(549-558).

Applied Data Stewardship

4 G‘s

“gather” = refers to all data ingestion, including the collecting and integrating of data assets through various systems of acquisition, as well as the policy instruments and ethics-based legislative frameworks through which the agency gains access to data and information (Rancourt, 2019). Sound data stewardship ensures that this data is acquired efficiently, ethically, and without duplication or redundancy.

Rancourt, E. (2019). The scientific approach as a transparency enabler throughout the data life-cycle. *Statistical Journal of the IAOS,* 35(549-558).

Applied Data Stewardship

4 G‘s

“guard” data = special attention is paid to access rights and privileges, data audit trails are performed, data monitoring and back-up protocols are systematized and ongoing, and metadata standards and classification systems are consistently updated (Rancourt, 2019). The goal in guarding data is to adhere to the “privacy by design” principles, ensuring that data is secure and encrypted, confidential and de-identified, and with all necessary privacy protocols in place in order to function ethically and according to our trust framework.

Rancourt, E. (2019). The scientific approach as a transparency enabler throughout the data life-cycle. *Statistical Journal of the IAOS,* 35(549-558).

Applied Data Stewardship

4 G‘s

“grow” data = the data is organized, processed, transformed, integrated, and extracted from for various uses (Rancourt, 2019). During this phase, data is cleaned and verified, quality assurance is performed, data is analyzed, explanations are developed, and hypotheses are tested. Efforts are made to grow data by ensuring its optimization and adhering to (and continually developing) data quality frameworks.

Rancourt, E. (2019). The scientific approach as a transparency enabler throughout the data life-cycle. *Statistical Journal of the IAOS,* 35(549-558).

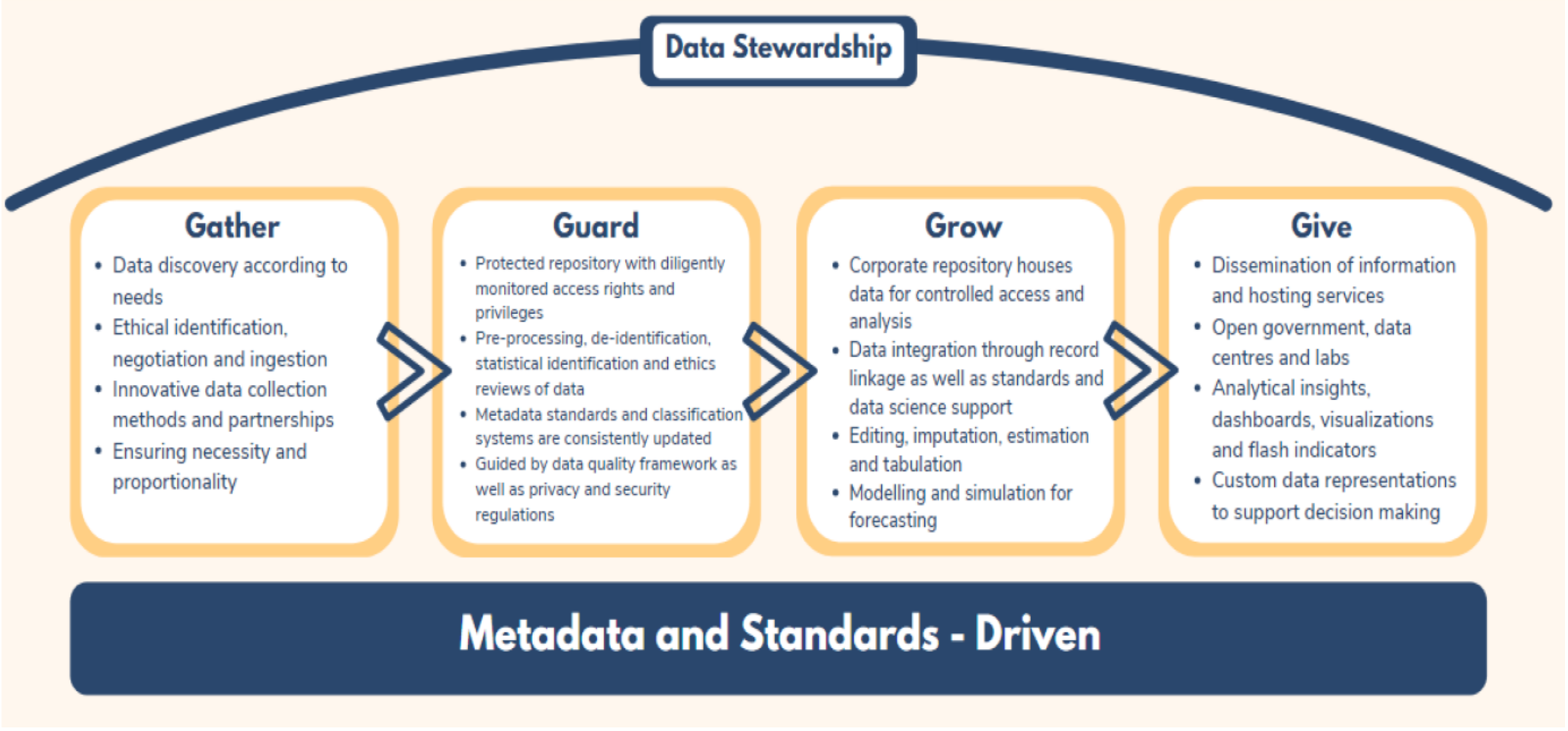
Applied Data Stewardship

4 G‘s

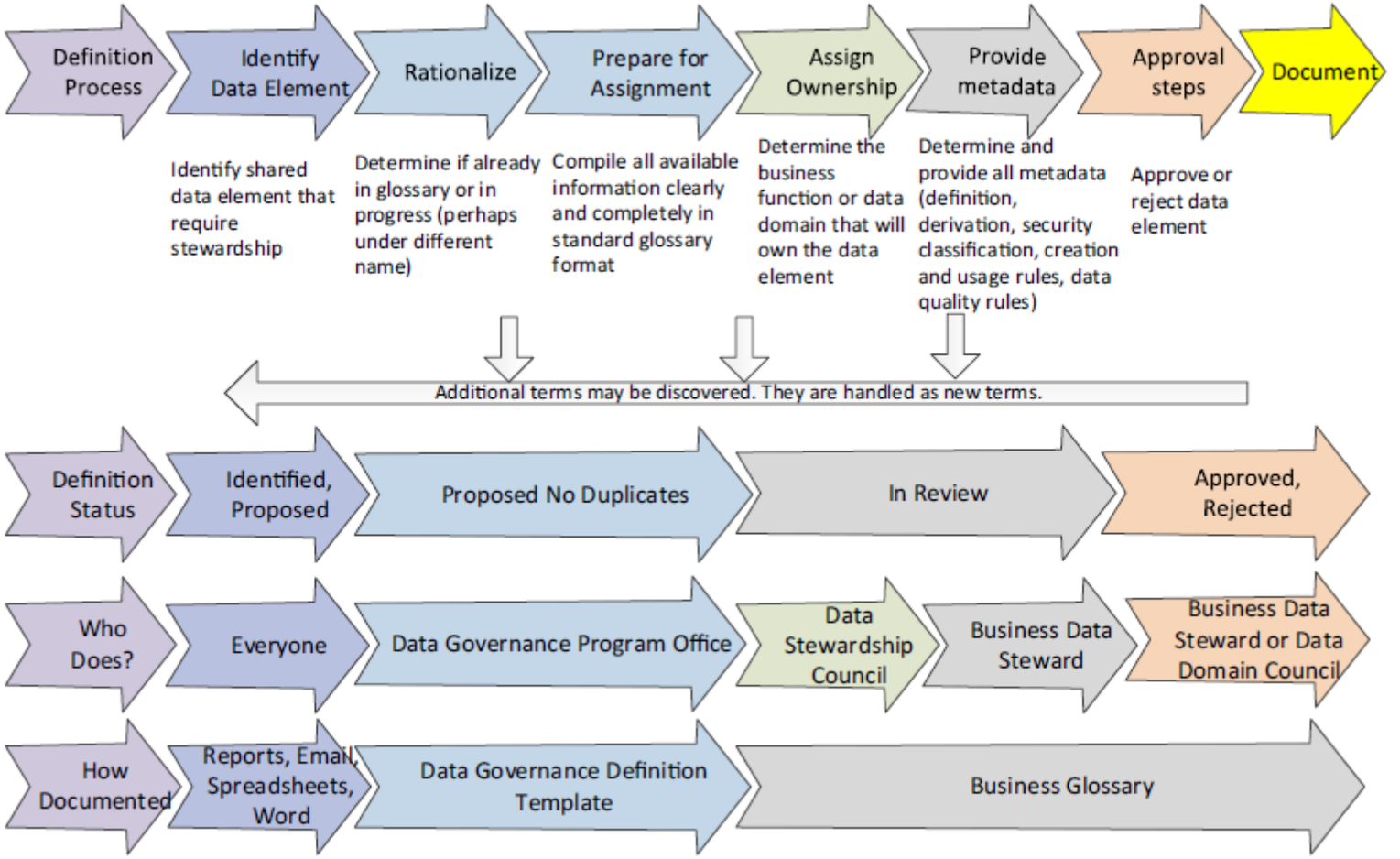
“give” = data and statistics are shared and published. Here, data access and interoperability are ensured, dissemination occurs regularly and with quality and accessibility, and the appropriate metadata is made available based on strategic requirements (Rancourt, 2019). The goal for Statistics Canada, and the public service more broadly, is to increase data discoverability and be “open by design”, having sharable and open data, metadata, metainformation, and analysis.

Rancourt, E. (2019). The scientific approach as a transparency enabler throughout the data life-cycle. *Statistical Journal of the IAOS,* 35(549-558).

Applied Data Stewardship



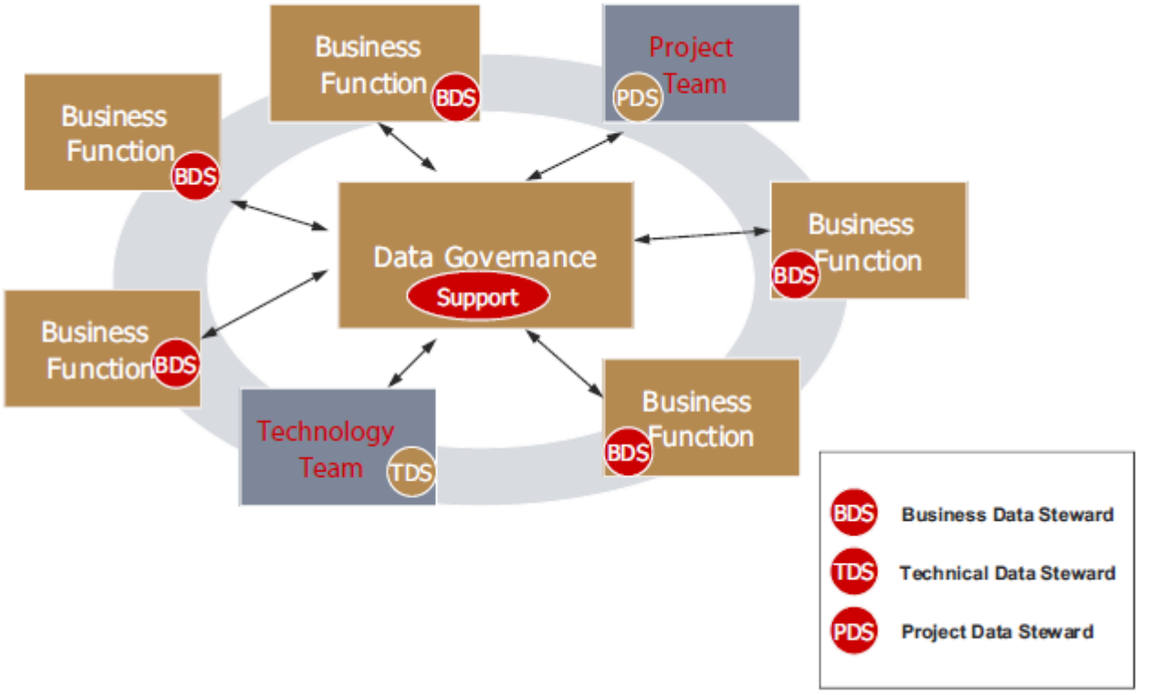
Starting Processes



The flow (across the top) of the process to identify, assign business function owner, define, and approve a new business data element.

Plotkin, D. (2021). Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance (2nd Ed.). London, UK: Academic Press.

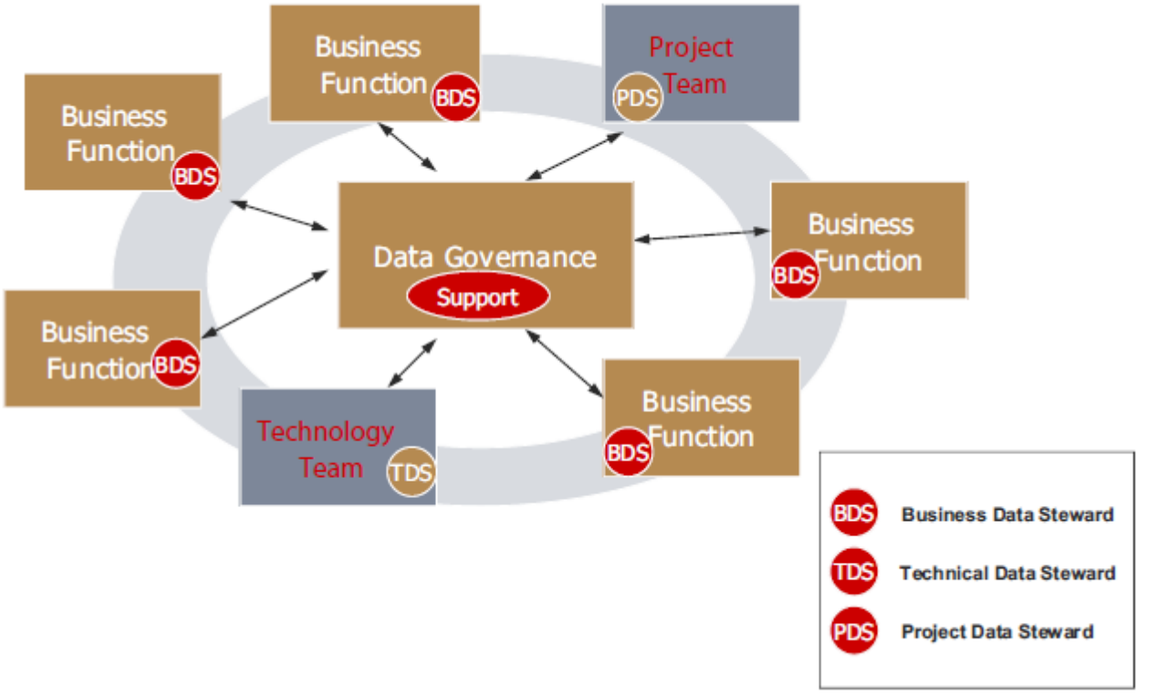
Starting Processes



The hybrid model of data stewardship on a project.

Plotkin, D. (2021). Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance (2nd Ed.). London, UK: Academic Press.

Starting Processes



The hybrid model of data stewardship on a project.

Plotkin, D. (2021). Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance (2nd Ed.). London, UK: Academic Press.

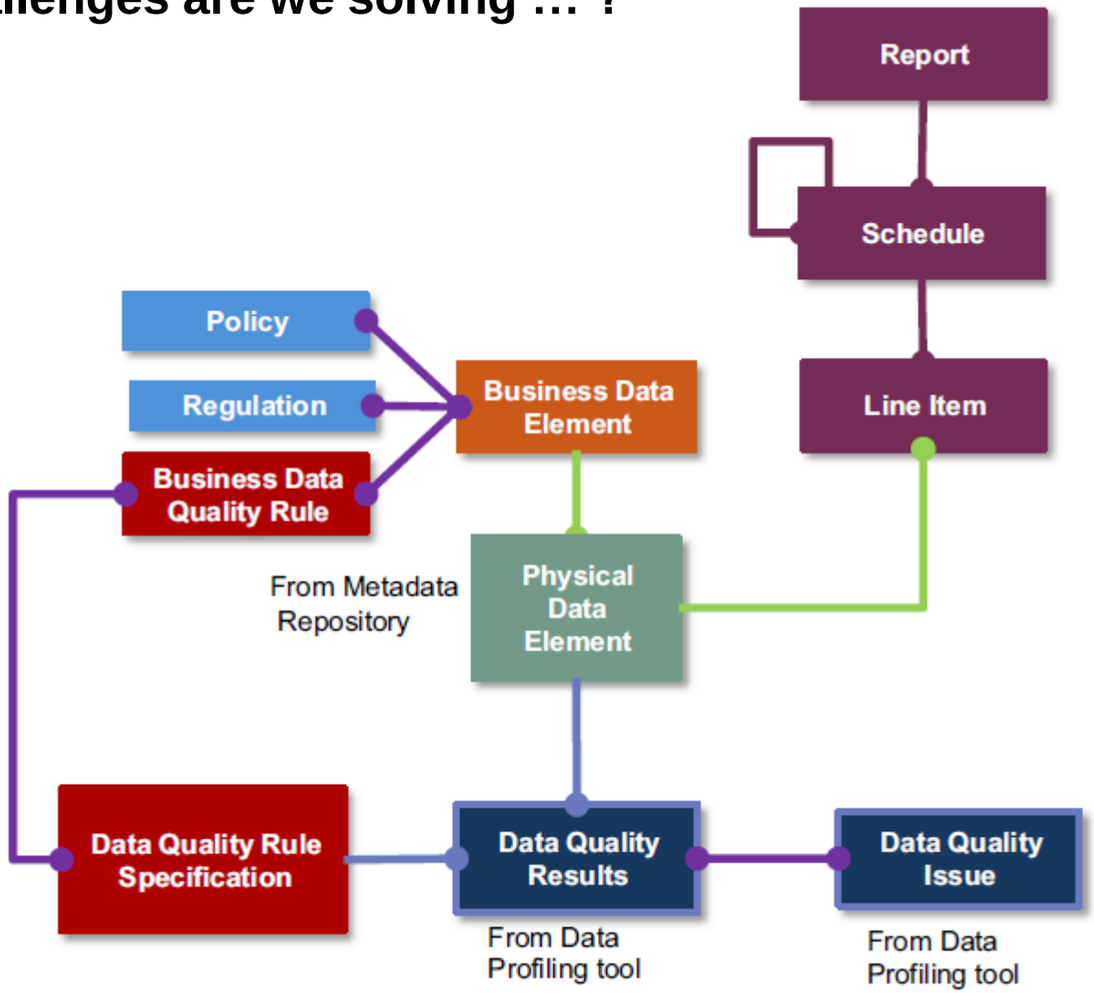
What challenges are we solving … ?

Plotkin, D. (2021). Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance (2nd Ed.). London, UK: Academic Press.

Table 6.6 Rationalizing the Data Elements

Data Element	Different Names or Different Data?	Total Data Points
Entry time	Ticket issue time, time of entry, transaction start time	1 or 4?
Prepayment time	Ticket paid time, payment time	1 or 3?
Amount due	Transaction total, transaction amount	1 or 3?
Payment method	Payment type	1 or 2?
Amount tendered	Amount paid, collected amount	1 or 3?
Change issued	Overpayment amount, refund amount, amount due to client	1 or 4?
Receipt issued	Receipt requested, receipt printed	1 or 3?
Actual exit time	Exit time, departure time	1 or 3?
Total		8 or 25?

What challenges are we solving … ?



Plotkin, D. (2021). Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance (2nd Ed.). London, UK: Academic Press.

Data Management Plan

https://datamanagement.hms.harvard.edu/training-events/rdm-seminar-series

Learning Goals

- ❑ Understand why data management plans are important
- ❑ Review example data management plan language
- ❑ Review DMPTool platform and templates

Data Sharing // Snafu in 3 Short Acts

https://youtu.be/66oNv_DJuPc

Research Data Management

"The **active and ongoing** management of data **through its lifecycle** of interest and usefulness to scholarship, science, and education."

—The University of Illinois' Graduate School of Library and Information Science

University of Applie
School of Life Scien

ACCESS & REUSE: Ensuring the broad utility of your research data efforts for other researchers

SHARE & DISSEMINATE: Esta blishing and supporting the reach and impact of your data

EVALUATE & ARCHIVE: Identify essential research records and evaluate for retention

STORE & MANAGE: Each stage of the Biomedical Data Lifecycle revolves around the management of data storage

PLAN & DESIGN: Plan processes from onboarding to project closure and data resources

COLLECT & CREATE: Organization and integration of data sets and collection processes

ANALYZE & COLLABORATE: Proc essing and analyzing data should be collaborative and documented

[Research Data Lifecycle by LMA RDMWG]

Research Planning

Concerns all aspects of preparing for a project:

- Seeking funding, awareness of Company and sponsor requirements
- Organization of data, records, tools, and/or resources needed to conduct the research and disseminate and archive valuable results

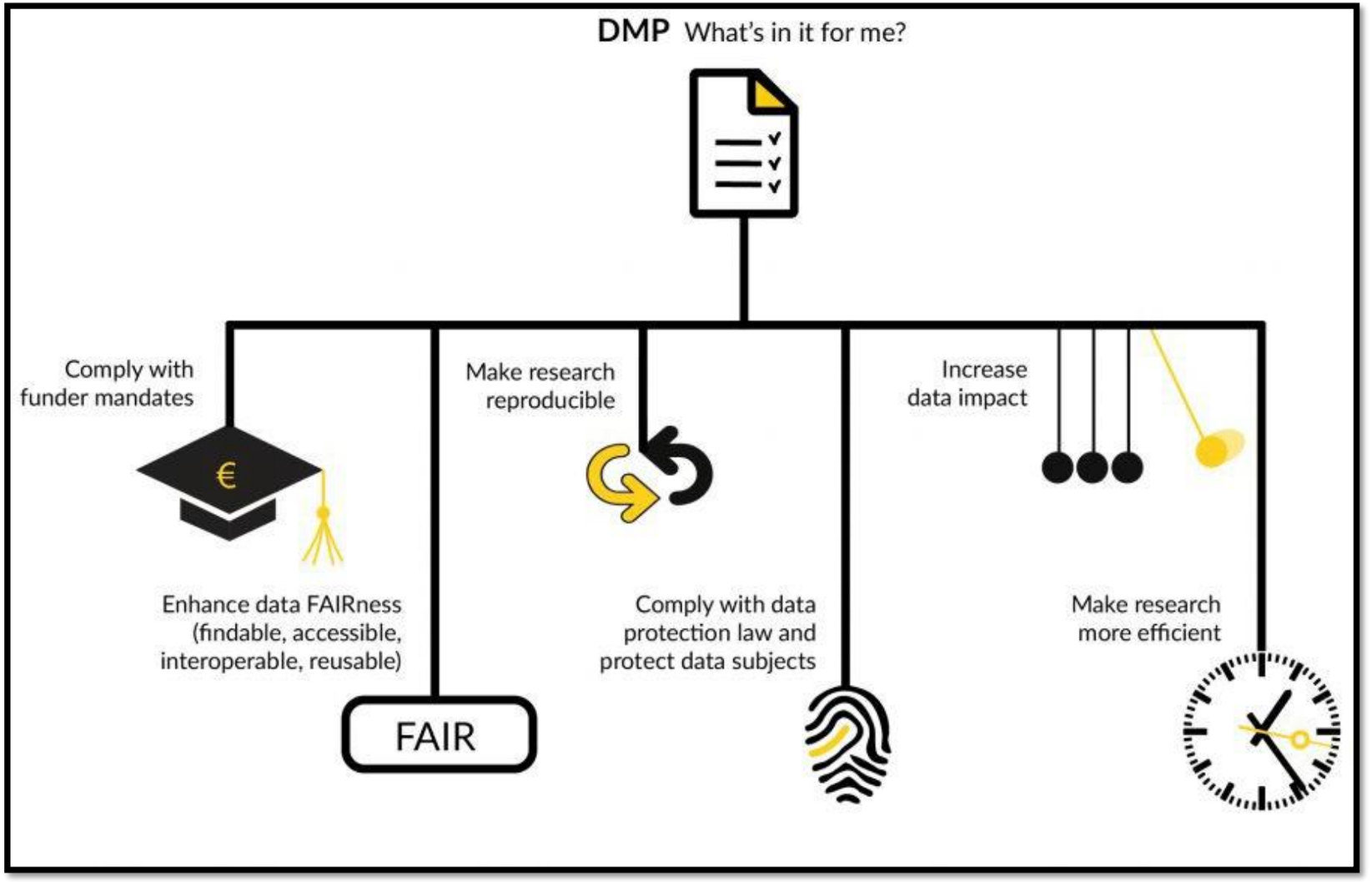
Important steps for sensitive data:

- If you know you will share data, get **permission** from participants as part of informed consent
- If you don't know if you'll share data in the future, consider getting permission anyway!

Data Management – Key Points

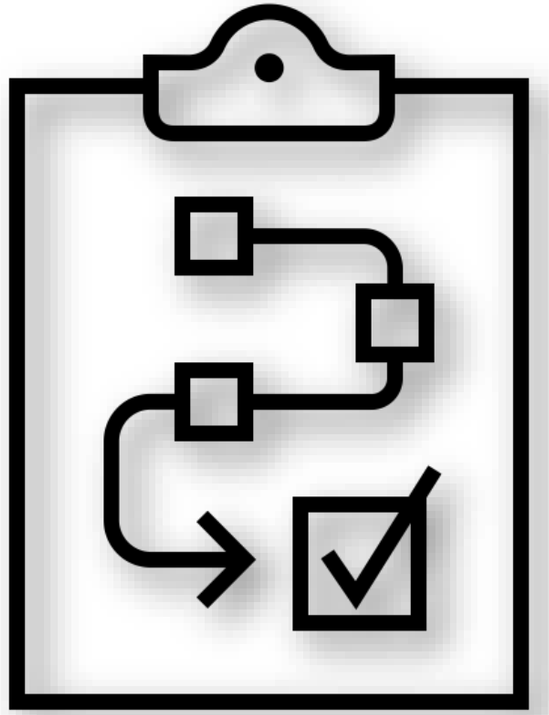
- ✓ Easier to analyze organized, documented data
- ✓ Find data more easily
- ✓ Don’t drown in irrelevant data
- ✓ Don’t lose data
- ✓ Get credit for your data
- ✓ Avoid accusations of misconduct

Research Planning



Data Management Plan

- ✓ Written document outlining plans for handling all of the data resulting from a research project in the short and long term
- ✓ Detailed procedures for data collection, all aspects of organization and processing before your data leaves your lab
- ✓ Plan for when data leave your lab so that others can find and access it in perpetuity
- ✓ Living, working document describing the entire project and is frequently referred to and updated



Poll Question

Have you written a Data Management Plan before?

https://www.menti.com/alqqhhxm6p3o

The voting code is **87 25 63 1**

Data Management and Sharing Plan (DMSP)

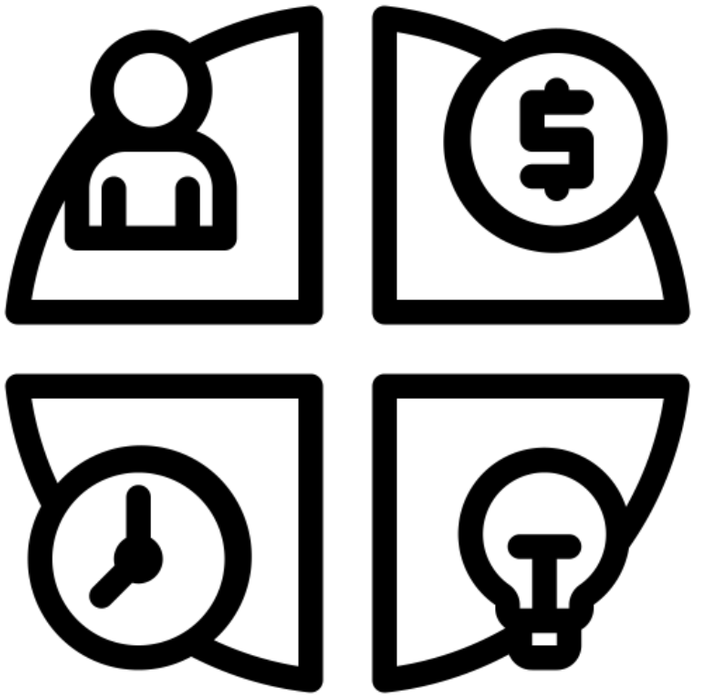
Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project.

https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

Data Management and Sharing Plan (DMSP)

<u>Required Elements*:</u>

1. Data Type
2. Related Tools, Software and/or Code
3. Standards
4. Data Preservation, Access, and Associated Timelines
5. Access, Distribution, or Reuse Considerations
6. Oversight of Data Management and Sharing



Data Management and Sharing Plan (DMSP)

1. **Data Type**
   - Identifying estimated type and amount of data to be generated (i.e., modality, level of aggregation, and degree of data processing)
   - Which data to be preserved and shared
   - Accompanying metadata, other relevant data, and associated documentation to be made available
2. **Related Tools, Software and/or Code**
   - Tools and software needed to access and manipulate data
3. **Standards**
   - Standards to be applied to scientific data and metadata

Data Management and Sharing Plan (DMSP)

4. **Data Preservation, Access, and Associated Timelines**
   - Proposed repository to be used consistent with Supplemental Information
   - How data will be findable and accessible (e.g., persistent unique identifier)
   - When data will be made available and for how long
5. **Access, Distribution, or Reuse Considerations**
   - Description of factors potentially affecting data access, distribution, or reuse related to informed consent or privacy and confidentiality protections
   - Whether access to human data will be controlled
6. **Oversight of Data Management and Sharing**
   - Plan compliance will be monitored/ managed and by whom

Example: Standards

Formal standards for our data have not yet been widely adopted. However, our data and other materials will be structured and described according to best practices.

STANDARDS

Data will be stored in common and open formats, such as **.txt files for our clinical data**. Information needed to make use of this **data will be recorded in codebooks** that will be accessible to the research team and will subsequently be shared alongside final datasets.

FORMATS

Information about our research process, including the details of our analysis pipeline will be **maintained contemporaneously, using protocols**. This information will be accessible to all members of the research team and will be shared alongside our data.

DOCUMENTATION

Example: Data Preservation, Access, & Associated Timelines

PRESERVATION

Data will be publicly shared and preserved in the repository **Zenodo**.

ACCESS

Zenodo is freely available to anyone to use, and the data will be securely stored in the CERN Data Center. Once deposited in Zenodo and published, the data set will be **assigned a DOI** and will be findable via a web search, the Zenodo repository search feature, or by the assigned DOI.

TIMELINES

The data will be published **concurrently with the first associated publication**, but no later than the end of the award period. The data will **be available for at least 20 years**; Zenodo's policies currently state "Items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least."

Example: Access, Distribution, or Reuse Considerations

In order **to ensure participant consent for data sharing**, IRB paperwork and informed consent documents will include language describing plans for data management and sharing data, describing the motivation for sharing, and explaining that **personal identifying information will be removed**.

CONSENT

ACCESS

To protect participant privacy and confidentiality, **shared data will be de-identified using the Safe Harbor method** per §164.514(b) to comply with the HIPAA Privacy Act.

DISTRIBUTION

Poll Question

Do you have a better understanding about Data Management Plans?

https://www.menti.com/alajov8q5oyu

The voting code is **25 69 85**

DMP Tool

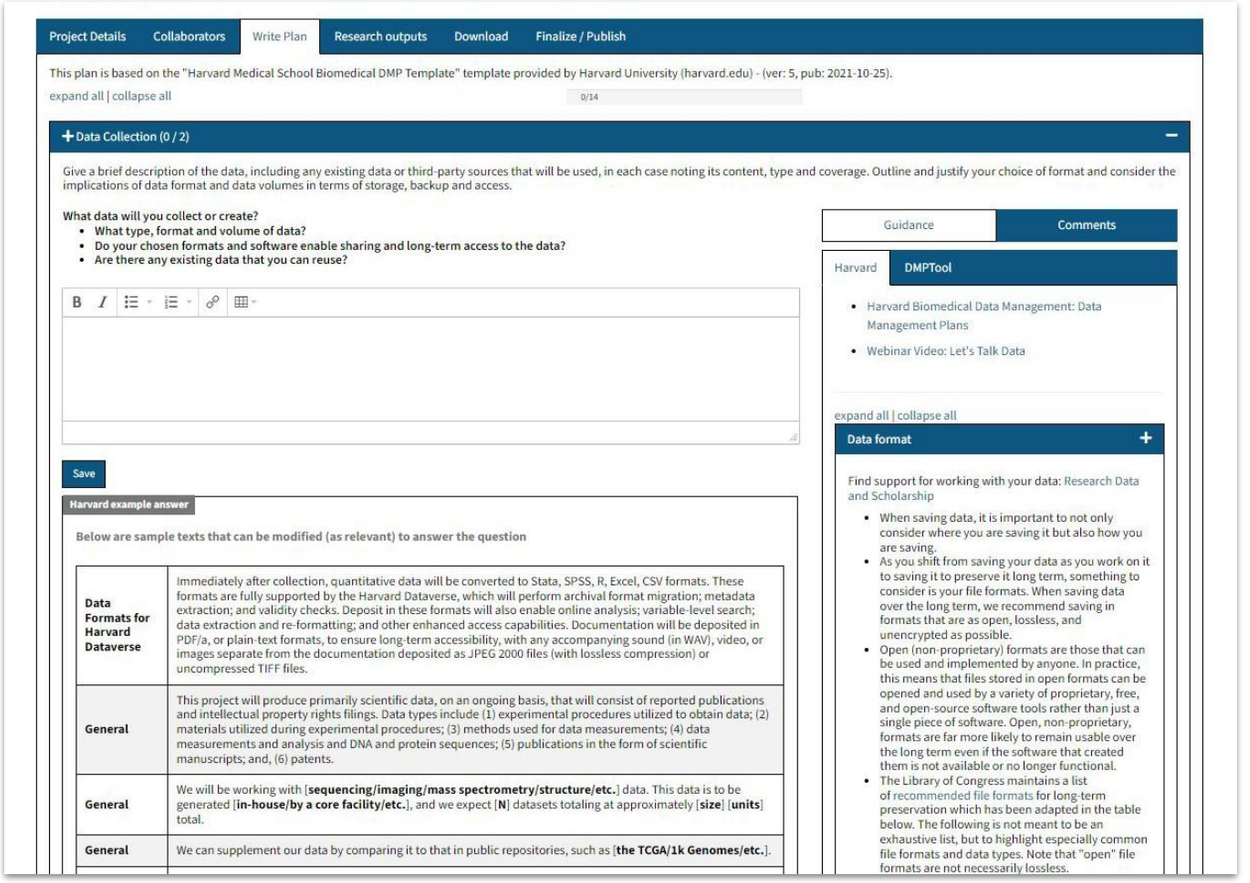
Learning Goals

- ❑ Review DMPTool platform
- ❑ Have knowledge about templates

DMSP Resources

- ❑ Customized **DMPTool** Templates
- ❑ Onboarding and Offboarding Checklists
- ❑ Data Use Agreement Resources
- ❑ ELN Resources (Electronic Lab Notebook)
- ❑ Metadata Guides
- ❑ Repository Resources

DMPTool Harvard Templates



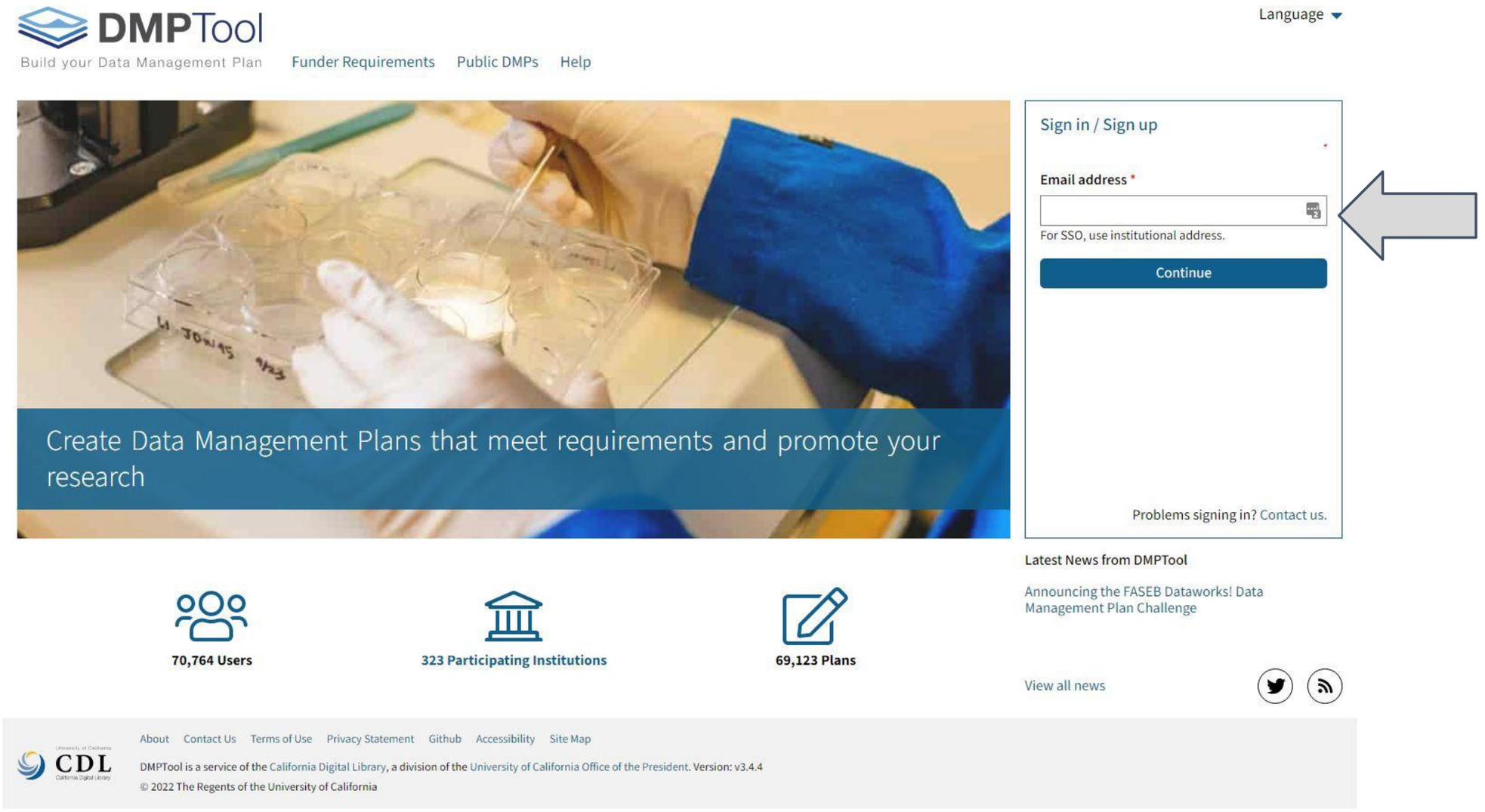
DMPTool
Build your Data Management Plan

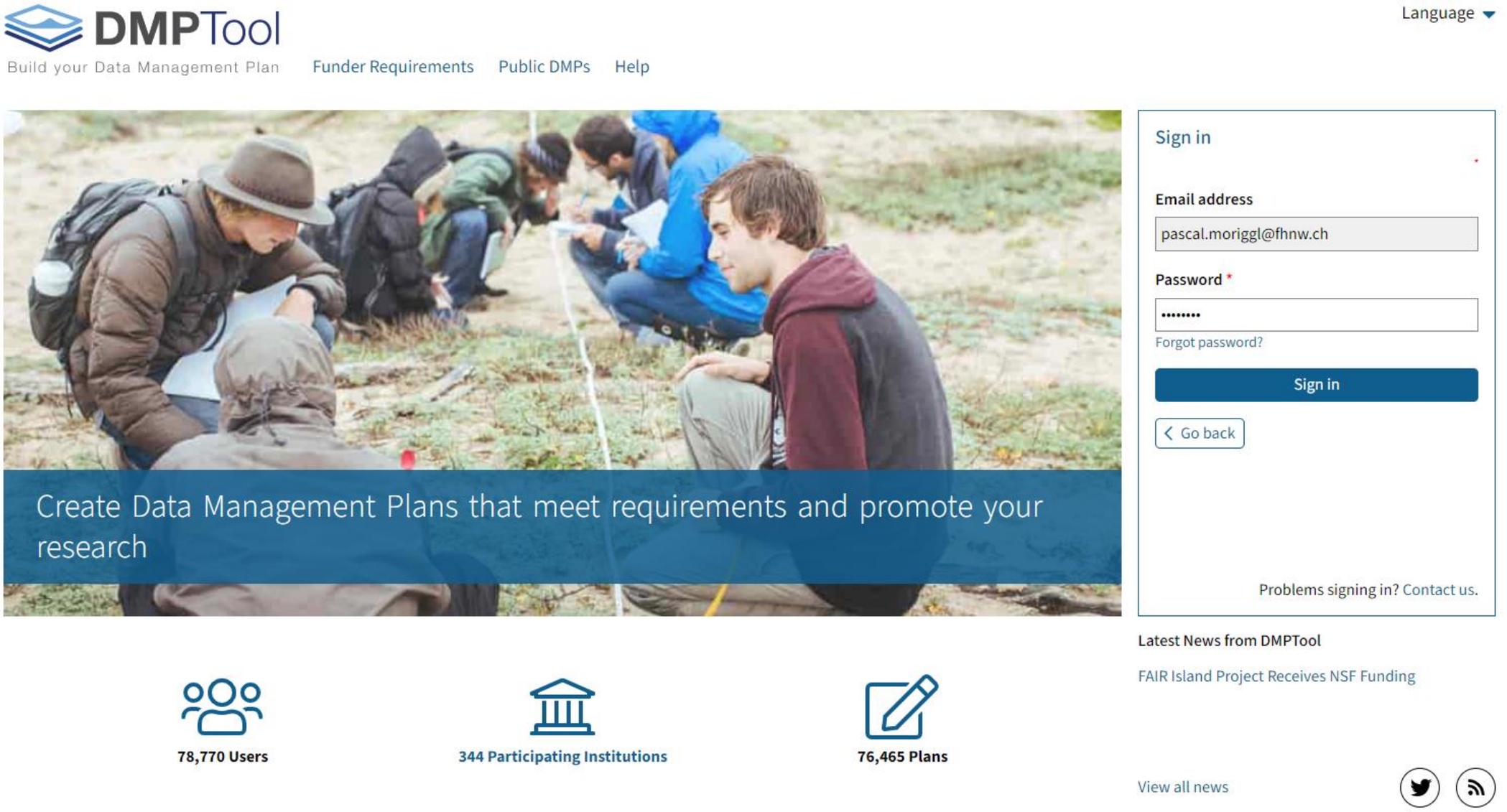
A free tool that helps researchers create data management plans that fulfill the requirements of different funders.

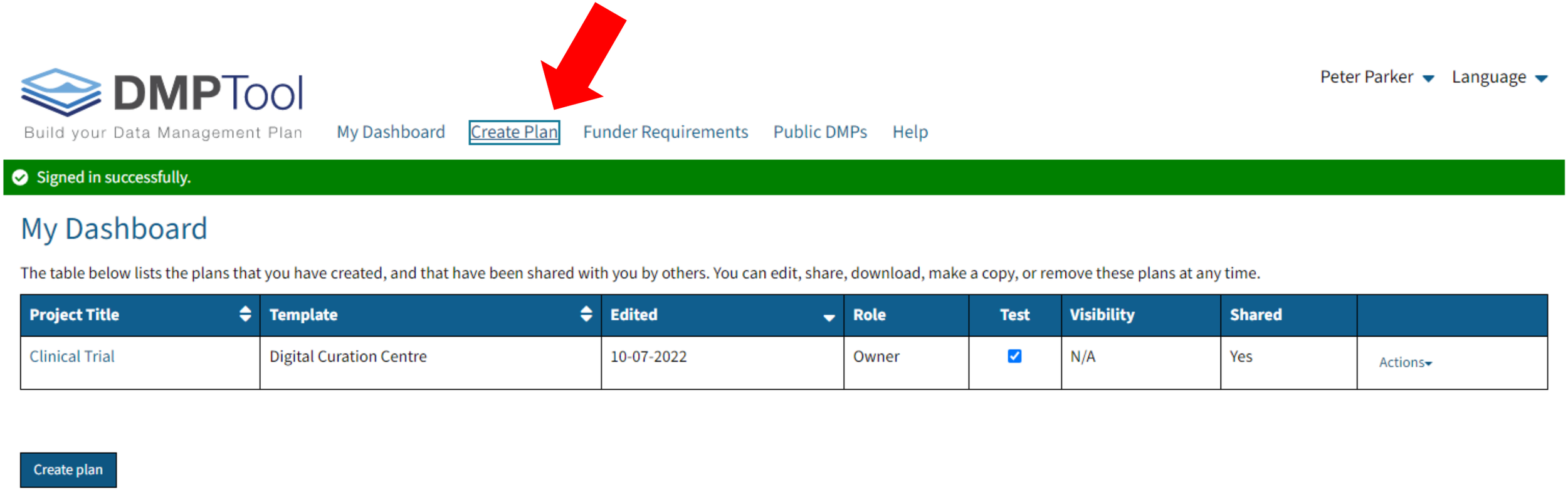
1. Go to https://dmptool.org
2. Sign up and login using your fhnw / zhaw mail address

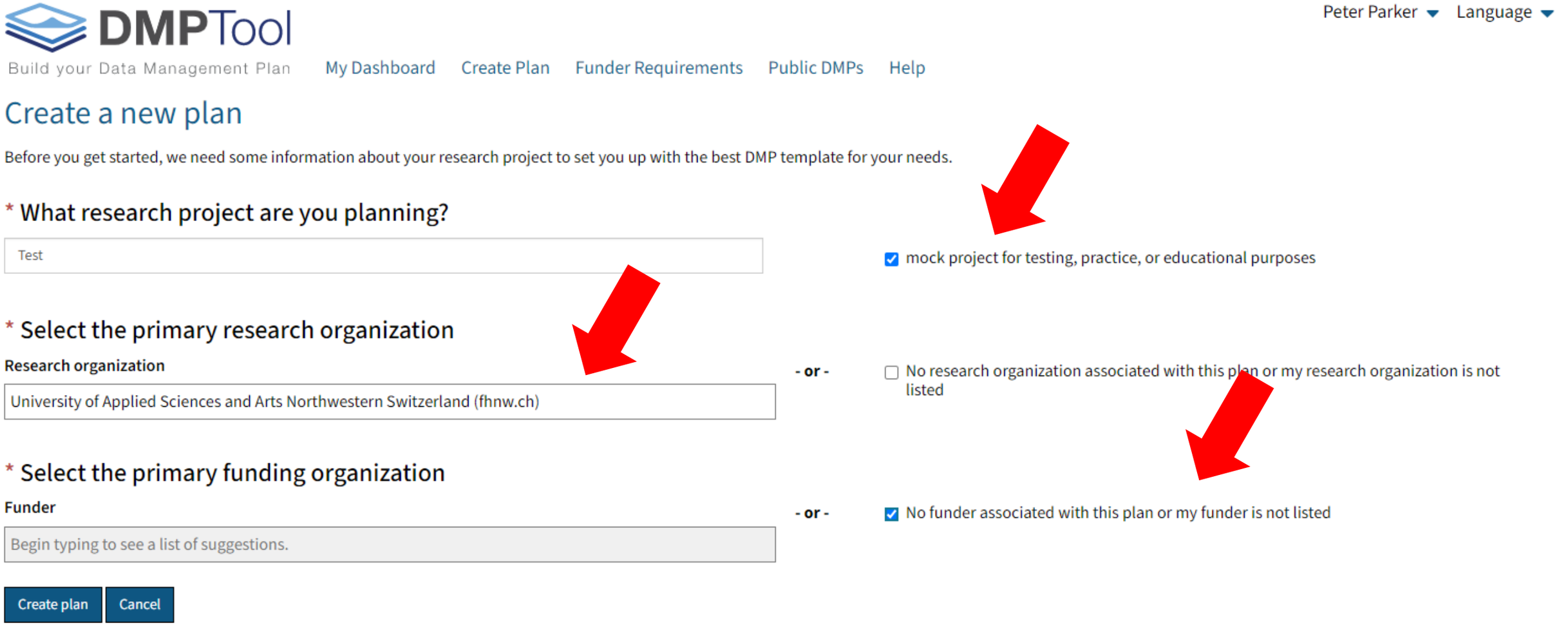
DMPTool Demo

https: /dmptool.org


DMPTool
Build your Data Management Plan
Funder Requirements
Public DMPs
Help
Language
Create Data Management Plans that meet requirements and promote your research
Sign in / Sign up
Email address *
For SSO, use institutional address.
Continue
Problems signing in? Contact us.
70,764 Users
323 Participating Institutions
69,123 Plans
Latest News from DMPTool
Announcing the FASEB Dataworks! Data Management Plan Challenge
View all news
About
Contact Us
Terms of Use
Privacy Statement
Github
Accessibility
Site Map
CDL
DMPTool is a service of the California Digital Library, a division of the University of California Office of the President. Version: v3.4.4
© 2022 The Regents of the University of California


DMPTool
Build your Data Management Plan
Funder Requirements
Public DMPs
Help
Language
Create Data Management Plans that meet requirements and promote your research
Sign in
Email address
pascal.moriggl@fhnw.ch
Password *
Forgot password?
Sign in
Go back
Problems signing in? Contact us.
78,770 Users
344 Participating Institutions
76,465 Plans
Latest News from DMPTool
FAIR Island Project Receives NSF Funding
View all news


DMPTool
Build your Data Management Plan
My Dashboard
Create Plan
Funder Requirements
Public DMPs
Help
Peter Parker
Language
Signed in successfully.
My Dashboard
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
Edited
Role
Test
Visibility
Shared
Clinical Trial
Digital Curation Centre
10-07-2022
Owner
N/A
Yes
Actions
Create plan


DMPTool
Build your Data Management Plan
Peter Parker
Language
My Dashboard
Create Plan
Funder Requirements
Public DMPs
Help
Create a new plan
Before you get started, we need some information about your research project to set you up with the best DMP template for your needs.
* What research project are you planning?
Test
mock project for testing, practice, or educational purposes
* Select the primary research organization
Research organization
University of Applied Sciences and Arts Northwestern Switzerland (fhnw.ch)
- or -
No research organization associated with this plan or my research organization is not listed
* Select the primary funding organization
Funder
Begin typing to see a list of suggestions.
- or -
No funder associated with this plan or my funder is not listed
Create plan
Cancel

Successfully created the plan.
This plan is based on the default template.

Test

Project Details | Collaborators | Write Plan | Research outputs | Download | Finalize / Publish

Project title *

Test

mock project for testing, practice, or educational purposes

Project abstract

Research domain

- Please select one -

Project Start

tt.mm.jjjj

Project End

tt.mm.jjjj

Research outputs may have ethical concerns

Funder

Select Guidance

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

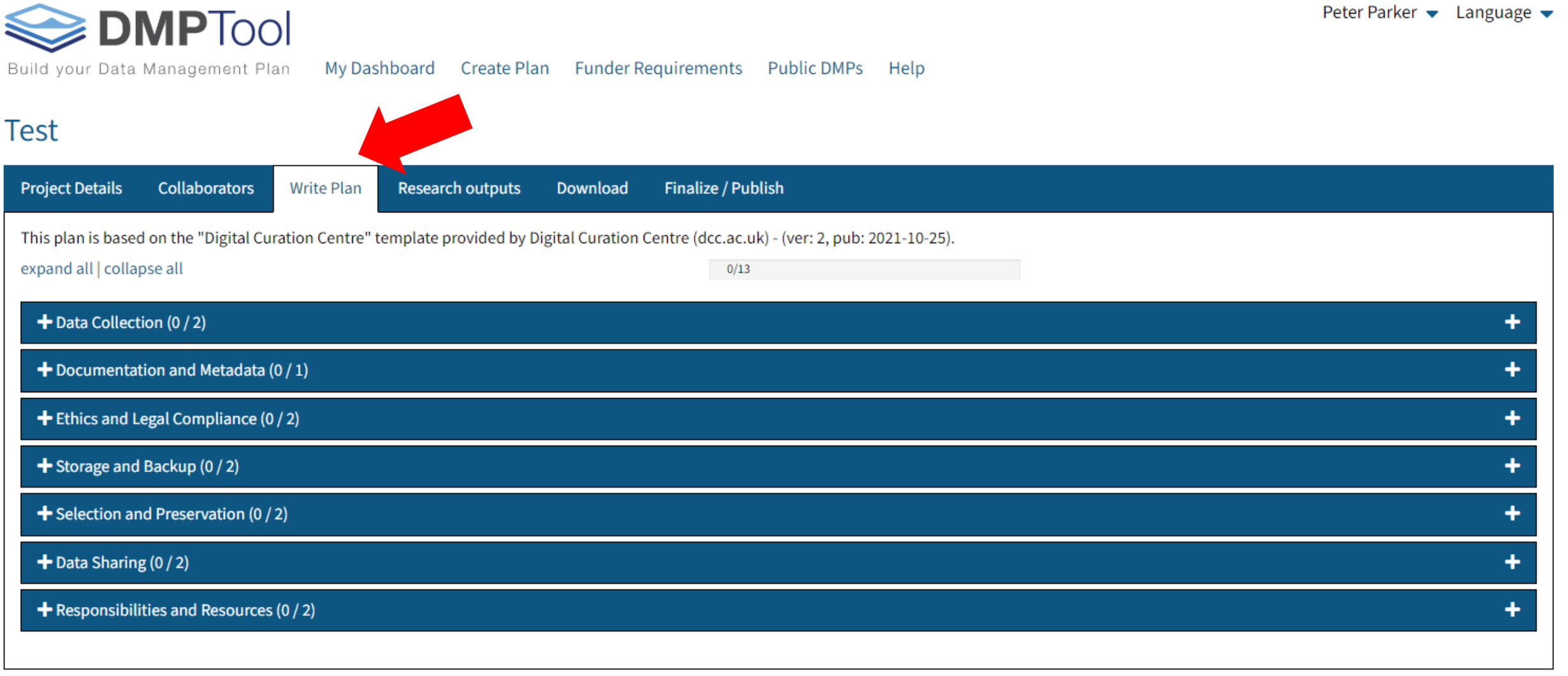
Select up to 6 organizations to see their guidance.

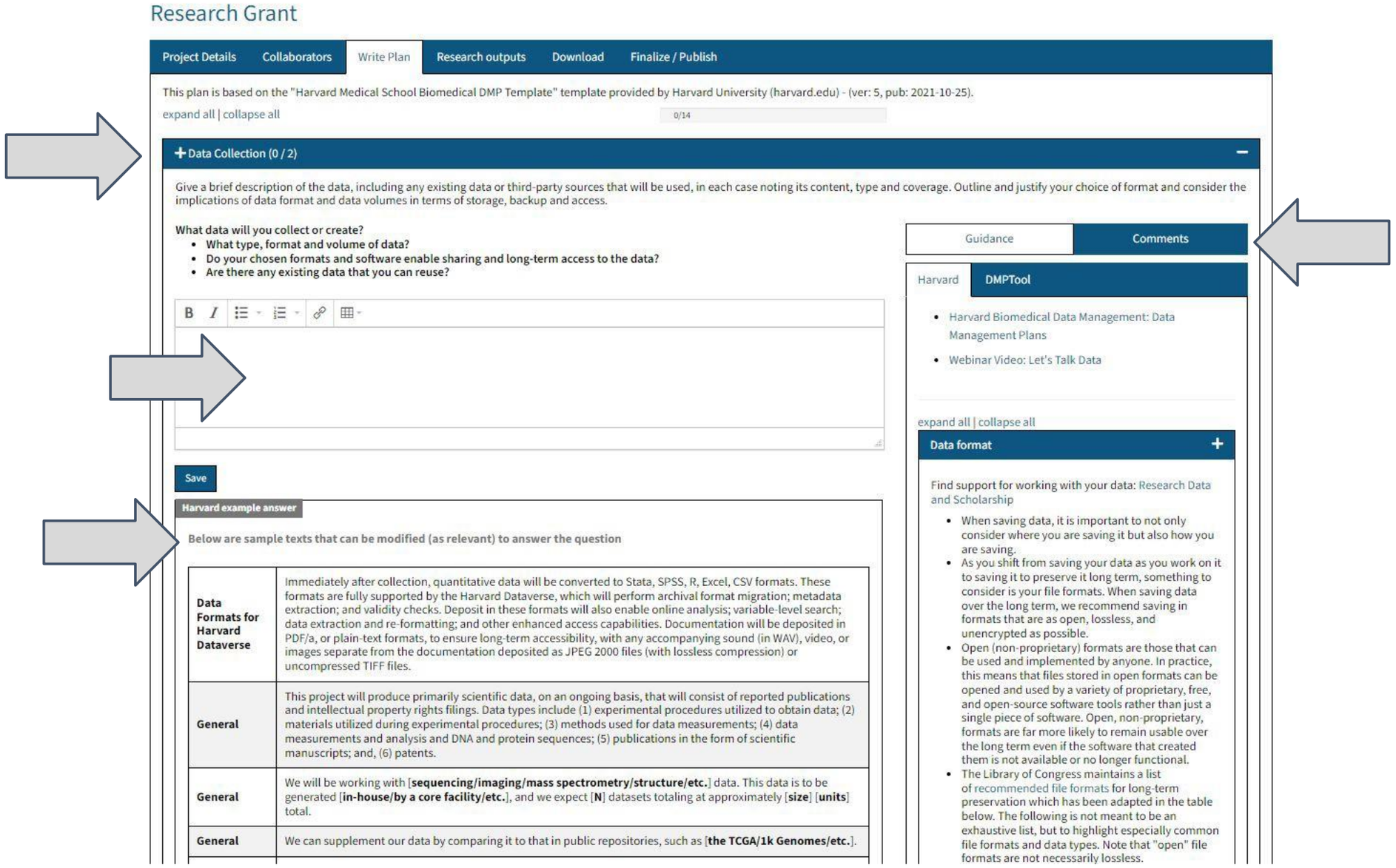
DMPTool

Find guidance from additional organizations below

See the full list

Save


DMPTool
Build your Data Management Plan
My Dashboard
Create Plan
Funder Requirements
Public DMPs
Help
Peter Parker
Language
Test
Project Details
Collaborators
Write Plan
Research outputs
Download
Finalize / Publish
This plan is based on the "Digital Curation Centre" template provided by Digital Curation Centre (dcc.ac.uk) - (ver: 2, pub: 2021-10-25).
expand all | collapse all
0/13
Data Collection (0 / 2)
Documentation and Metadata (0 / 1)
Ethics and Legal Compliance (0 / 2)
Storage and Backup (0 / 2)
Selection and Preservation (0 / 2)
Data Sharing (0 / 2)
Responsibilities and Resources (0 / 2)


Research Grant
Project Details
Collaborators
Write Plan
Research outputs
Download
Finalize / Publish
This plan is based on the "Harvard Medical School Biomedical DMP Template" template provided by Harvard University (harvard.edu) - (ver: 5, pub: 2021-10-25).
expand all | collapse all
0/14
+ Data Collection (0 / 2)
Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.
What data will you collect or create?
What type, format and volume of data?
Do your chosen formats and software enable sharing and long-term access to the data?
Are there any existing data that you can reuse?
Save
Harvard example answer
Below are sample texts that can be modified (as relevant) to answer the question
Data Formats for Harvard Dataverse
Immediately after collection, quantitative data will be converted to Stata, SPSS, R, Excel, CSV formats. These formats are fully supported by the Harvard Dataverse, which will perform archival format migration; metadata extraction; and validity checks. Deposit in these formats will also enable online analysis; variable-level search; data extraction and re-formatting; and other enhanced access capabilities. Documentation will be deposited in PDF/a, or plain-text formats, to ensure long-term accessibility, with any accompanying sound (in WAV), video, or images separate from the documentation deposited as JPEG 2000 files (with lossless compression) or uncompressed TIFF files.
General
This project will produce primarily scientific data, on an ongoing basis, that will consist of reported publications and intellectual property rights filings. Data types include (1) experimental procedures utilized to obtain data; (2) materials utilized during experimental procedures; (3) methods used for data measurements; (4) data measurements and analysis and DNA and protein sequences; (5) publications in the form of scientific manuscripts; and, (6) patents.
General
We will be working with [sequencing/imaging/mass spectrometry/structure/etc.] data. This data is to be generated [in-house/by a core facility/etc.], and we expect [N] datasets totaling at approximately [size] [units] total.
General
We can supplement our data by comparing it to that in public repositories, such as [the TCGA/1k Genomes/etc.].
Guidance
Comments
Harvard
DMPTool
Harvard Biomedical Data Management: Data Management Plans
Webinar Video: Let's Talk Data
expand all | collapse all
Data format
Find support for working with your data: Research Data and Scholarship
When saving data, it is important to not only consider where you are saving it but also how you are saving.
As you shift from saving your data as you work on it to saving it to preserve it long term, something to consider is your file formats. When saving data over the long term, we recommend saving in formats that are as open, lossless, and unencrypted as possible.
Open (non-proprietary) formats are those that can be used and implemented by anyone. In practice, this means that files stored in open formats can be opened and used by a variety of proprietary, free, and open-source software tools rather than just a single piece of software. Open, non-proprietary, formats are far more likely to remain usable over the long term even if the software that created them is not available or no longer functional.
The Library of Congress maintains a list of recommended file formats for long-term preservation which has been adapted in the table below. The following is not meant to be an exhaustive list, but to highlight especially common file formats and data types. Note that "open" file formats are not necessarily lossless.

Key takeaways

- Data management is important to keep data safe from harm and make data usable and discoverable
- A data management plan includes strategies and processes to organize, describe, preserve, and share data
- Some funding agencies have their own DMP policy
- DMPTool can be used for templates and guidance to create a DMP

Further Reading

Borghi J, Abrams S, Lowenberg D, Simms S, Chodacki J (2018) Support Your Data: A Research Data Management Guide for Researchers. Research Ideas and Outcomes 4: e26439. https://doi.org/10.3897/rio.4.e26439

Briney KA, Coates H, Goben A (2020) Foundational Practices of Research Data Management. Research Ideas and Outcomes 6: e56508. https://doi.org/10.3897/rio.6.e56508

Corpas M, Kovalevskaya NV, McMurray A, Nielsen FGG (2018) A FAIR guide for data providers to maximise sharing of human genomic data. PLoS Comput Biol 14(3): e1005873. https://doi.org/10.1371/journal.pcbi.1005873

Wilkinson, M, Dumontier, M, Aalbersberg, I, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3: 160018. https://doi.org/10.1038/sdata.2016.18

Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, Teal TK (2017) Good enough practices in scientific computing. PLoS Comput Biol 13(6): e1005510. https://doi.org/10.1371/journal.pcbi.1005510