

eu:openscreen



 **Fraunhofer**
ITMP

A Comprehensive Workshop on Preclinical Drug Discovery (CDD)

Data Management




Andrea Zaliani Fraunhofer ITMP

Supported by:



Volkswagen**Stiftung**

Outline of presentation

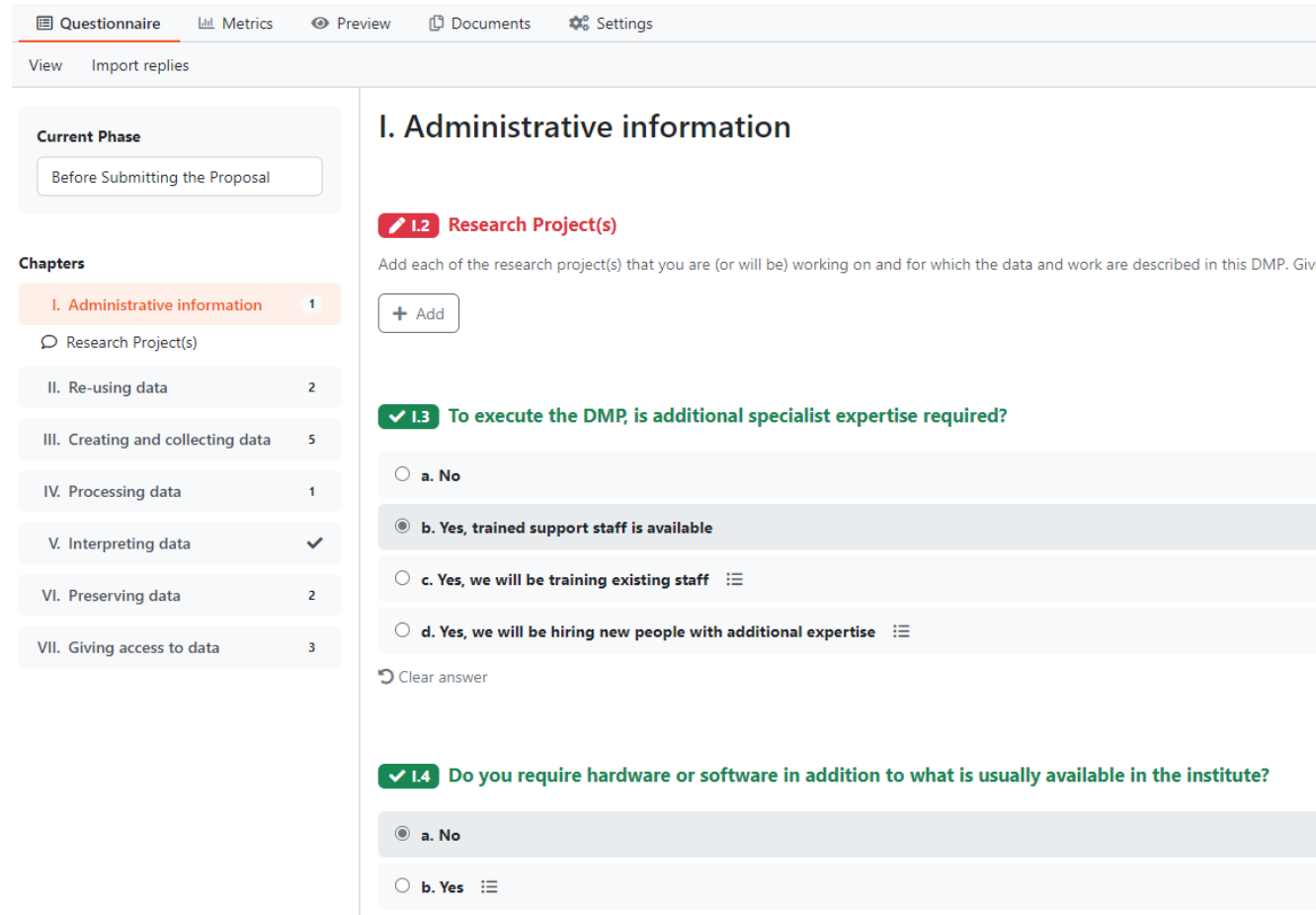
- Data Types
- Data Management Core 
- Example
- Data maturity assessment 
- Example from EU Pharma company + Exercise 
- Summary

Data types

- In principle, they can be everything
- In practice, they need to be:
 - Machine – readable
 - Standardized according to a recognizable format
- For Drug Discovery the most frequent are:
 - Documents (PDF, DOCX, TXT, CSV)
 - Tables (XLSX, CSV, DAT)
 - Pictures (PNG, JPG, TIFF, Depending on reader)
 - Molecules [protein, peptides, RNA/DNA, small organic]
 - (SMILES, InChi, SDF, MOL2, PDB)
 - Protein/Gene Sequences (TXT, FASTA, ClustalW.....)

Core of Data Management

- DMP is the core of data management effort
- How do we prepare it?
- What must it contain?

A screenshot of the DS-Wizard web application for creating a Data Management Plan (DMP). The interface has a top navigation bar with tabs: "Questionnaire" (active), "Metrics", "Preview", "Documents", and "Settings". Below this is a sub-bar with "View" and "Import replies". On the left, a sidebar shows the "Current Phase" as "Before Submitting the Proposal" and a list of "Chapters" with progress indicators: I. Administrative information (1), Research Project(s), II. Re-using data (2), III. Creating and collecting data (5), IV. Processing data (1), V. Interpreting data (checked), VI. Preserving data (2), and VII. Giving access to data (3). The main content area is titled "I. Administrative information" and contains two sections. The first, "1.2 Research Project(s)", includes a text prompt and a "+ Add" button. The second, "1.3 To execute the DMP, is additional specialist expertise required?", has four radio button options: "a. No", "b. Yes, trained support staff is available" (selected), "c. Yes, we will be training existing staff", and "d. Yes, we will be hiring new people with additional expertise". Below this is a "Clear answer" link. The third section, "1.4 Do you require hardware or software in addition to what is usually available in the institute?", has two radio button options: "a. No" (selected) and "b. Yes".



Practical example on how to build a DMP

- <https://researchers.ds-wizard.org/wizard/>

FAIR concept

- F = Findable
- A = Accessible
- I = Interoperable
- R = Re-usable


Or browse the various sections

<p>F Findability</p> <p>EXEMPLAR RECIPES</p> <ul style="list-style-type: none">• Unique, persistent identifiers• Search engine optimization <p>LEARN MORE</p>	<p>A Accessibility</p> <p>EXEMPLAR RECIPES</p> <ul style="list-style-type: none">• Transferring data with SFTP• Downloading data with Aspera <p>LEARN MORE</p>	<p>I Interoperability</p> <p>EXEMPLAR RECIPES</p> <ul style="list-style-type: none">• Selecting terminologies and ontologies• Creating a metadata profile <p>LEARN MORE</p>	<p>R Reusability</p> <p>EXEMPLAR RECIPES</p> <ul style="list-style-type: none">• Data licenses• Declaring data's permitted uses <p>LEARN MORE</p>
<p>Infrastructure</p> <p>LEARN MORE</p>	<p>Assessments</p> <p>LEARN MORE</p>	<p>Applied Examples</p> <p>LEARN MORE</p>	<p>Maturity model</p> <p>LEARN MORE</p>

FAIR data maturity levels explained

5	Managed Data Assets	Enterprise Level. Data at this level is optimally managed at the most granular level in an environment offering <i>data governance, master data management and reference data management</i> capabilities.
4	Semantically Typed Data	Cross-community Level. This level focuses on cross-domain interoperability and is meant to be the level required for larger harmonization and integration projects.
3	Standardised Data	Community Level. Data at this level complies with community standard domain models, terminologies and formats, and is hosted in an environment offering searching and retrieval capabilities.
2	Described Data	Project Level. All datasets generated within a project are consistently described against a locally defined schema, controlled terminologies, and hosted in an environment offering data catalogue level searching capabilities.
1	Identifiable Data	Data Object level. Data at this level is identifiable as individual generic data objects and described by generic metadata elements. Hosting environment offers <i>limited</i> retrieval capabilities.
0	Single Use Data	No potential for re-use beyond lifetime of the research project

Examples of dataset



Identifier	Name	GFP plate reader anti-viral screen EC50 (μM)	GFP plate reader anti-viral screen pEC50	HCI antiviral screen EC50 (μM)	HCI antiviral screen pEC50	ATPlite toxicity CC50 (μM)	ATPlite toxicity pCC50
COVC-0134244088	LAGOCHILIN	> 20.00	<4.7				
COVC-0134273678	SABCOMELINE HYDROCHLORIDE	> 20.00	<4.7				
COVC-0134286843	EVP-6124_ENCENICLINE	>100.00	<4	>100.00	<4	20.9	4.68
COVC-0134309031	RUPINTRIVIR	> 25.10; > 20.00	<4.6; <4.7				
COVC-0134421928	GDC-0980	> 20.00	<4.7				
COVC-0134421965	SOTRASTAUIN	> 20.00	<4.7				
COVC-0134436610	NGP-555	> 20.00	<4.7				
COVC-0134754952	XL-147	> 20.00	<4.7				
COVC-0134874627	CLOBETASOL	> 20.00	<4.7				
COVC-0134874669	CLOPERASTINE HYDROCHLORIDE	> 20.00	<4.7				
COVC-0134874682	T-62	> 20.00	<4.7				
COVC-0134874690	NIBENTAN	> 20.00	<4.7				
COVC-0134874742	IOFETAMINE HYDROCHLORIDE	> 20.00	<4.7				

Examples of dataset minimal optimized

Identifier	Synonyms	PubChem ID	Assay 1				
			GFP plate reader anti-viral screen	Type	Relation	Value	Unit
COVC-0134244088	LAGOCHILIN	2159	Text with description	EC50	>	20	μM
COVC-0134273678	SABCOMELINE HYDROCHLORIDE	65947	Text with description	EC50	>	20	μM
COVC-0134286843	EVP-6124_ENCENICLINE	9871419	Text with description	EC50	>	100	μM
COVC-0134309031	RUPINTRIVIR	7405	Text with description	EC50	>	25.1	μM
COVC-0134309032	RUPINTRIVIR	7405	Text with description	EC50	>	20.01	μM
COVC-0134421928	GDC-0980	4626	Text with description	EC50	>	20	μM
COVC-0134421965	SOTRASTAUIN	5281515	Text with description	EC50	>	20	μM
COVC-0134436610	NGP-555	65957	Text with description	EC50	>	20	μM
COVC-0134754952	XL-147	5479529	Text with description	EC50	>	20	μM
COVC-0134874627	CLOBETASOL	82148	Text with description	EC50	>	20	μM
COVC-0134874669	CLOPERASTINE HYDROCHLORIDE	156419	Text with description	EC50	>	20	μM
COVC-0134874682	T-62	5362065	Text with description	EC50	>	20	μM
COVC-0134874690	NIBENTAN	71188	Text with description	EC50	>	20	μM
COVC-0134874742	IOFETAMINE HYDROCHLORIDE	72172	Text with description	EC50	>	20	μM
COVC-0134874745	TU-2100	72402	Text with description	EC50	>	20	μM

Summary

- Any project you will start will require:

- Data from
 - past
 - to be produced
- Resources
 - People
 - Money
- Support



FAIR Assessment on reusability

DMP collects ALL of this info

Questions?

Supported by: