A Comprehensive Workshop on Preclinical Drug Discovery (CDD)

Data Management

Andrea Zaliani Fraunhofer ITMP

Supported by: VolkswagenStiftung

Outline of presentation

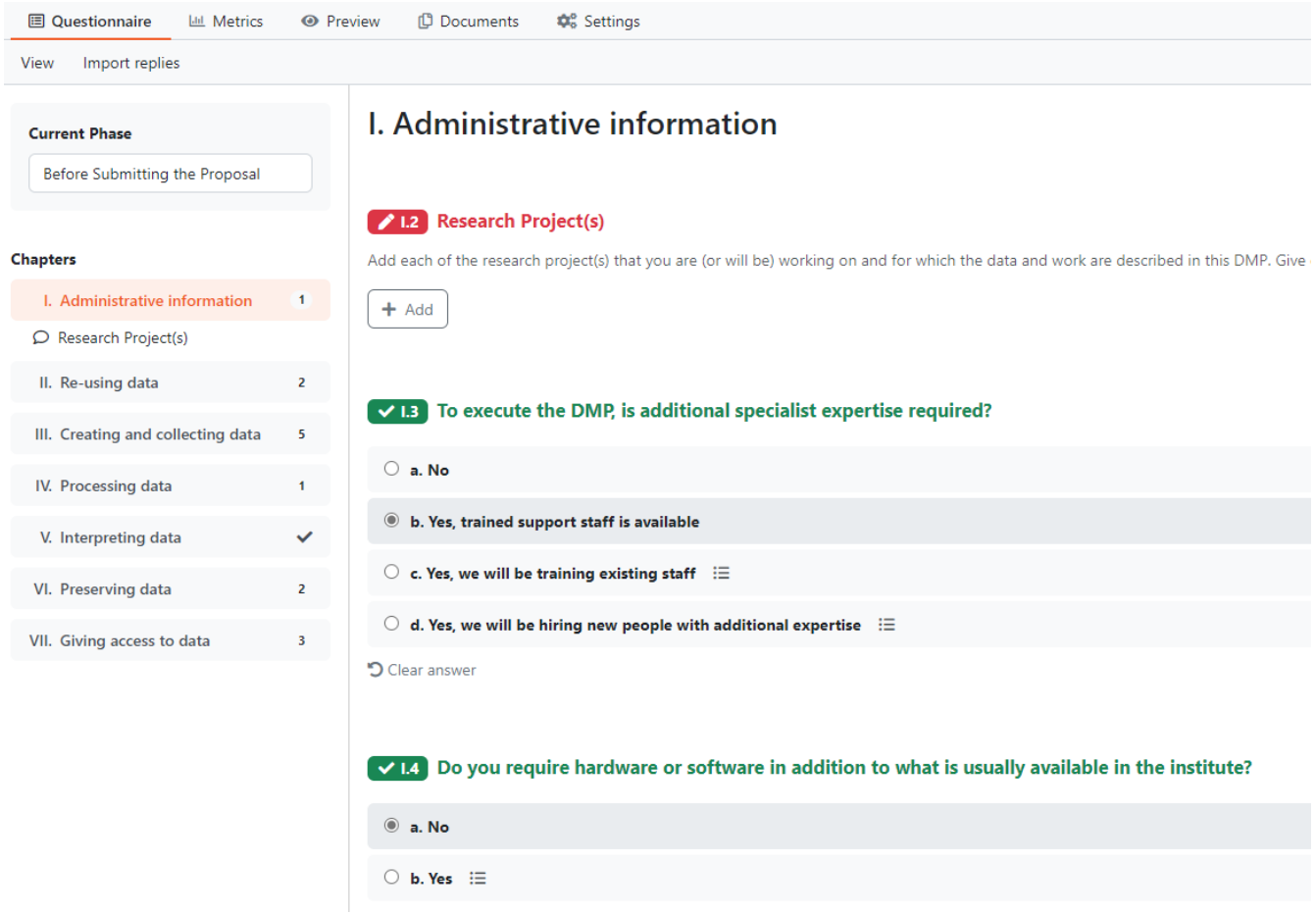
- Data Types
- Data Management Core
- Example
- Data maturity assessment
- Example from EU Pharma company + Exercise
- Summary

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?



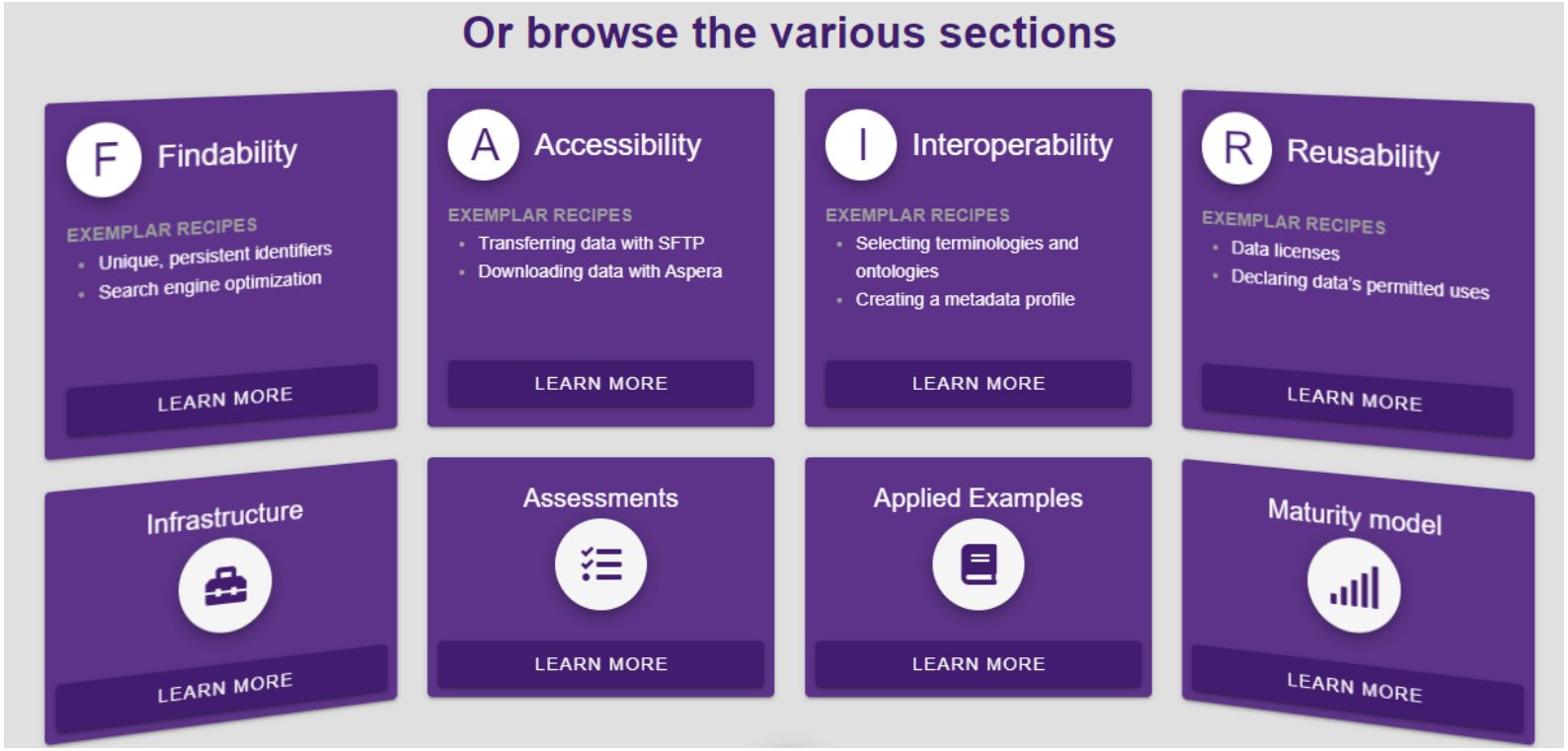
https://researchers.ds-wizard.org/wizard/

Practical example on how to build a DMP

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

FAIR concept

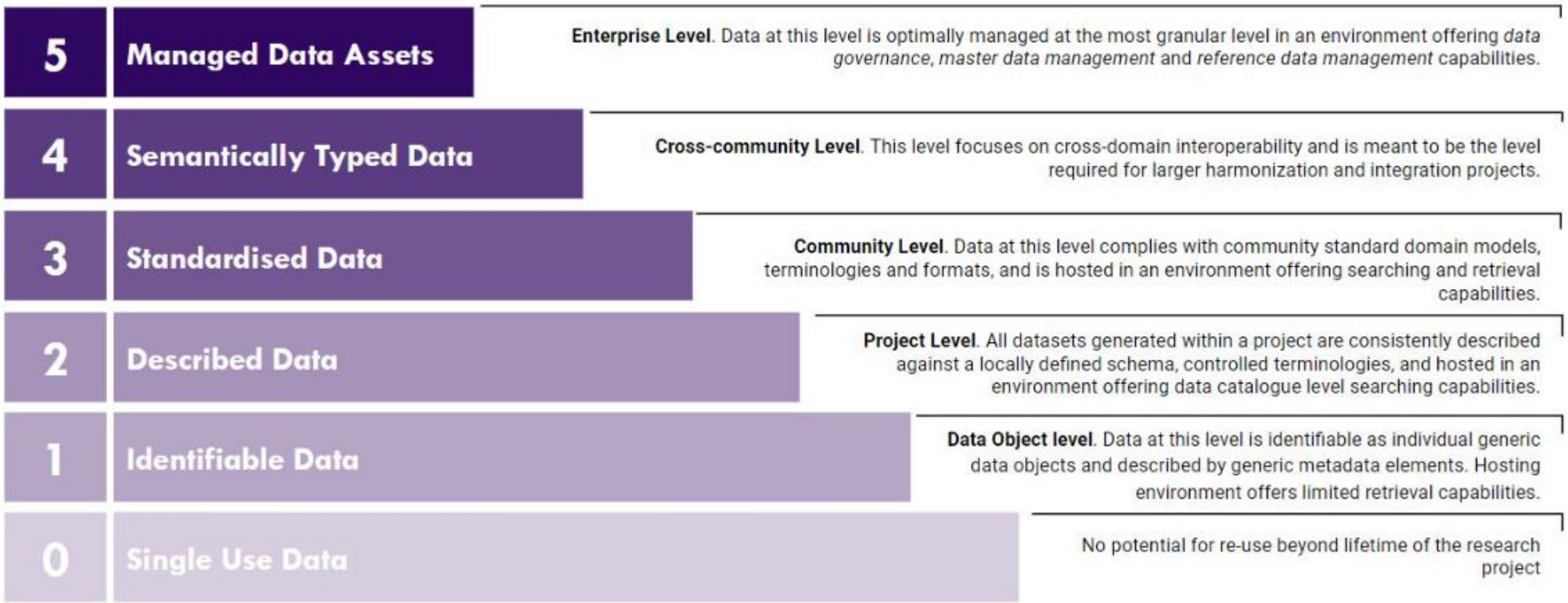
- F = Findable
- A = Accessible
- I = Interoperable
- R = Re-usable



https://faircookbook.elixir-europe.org/content/home.html

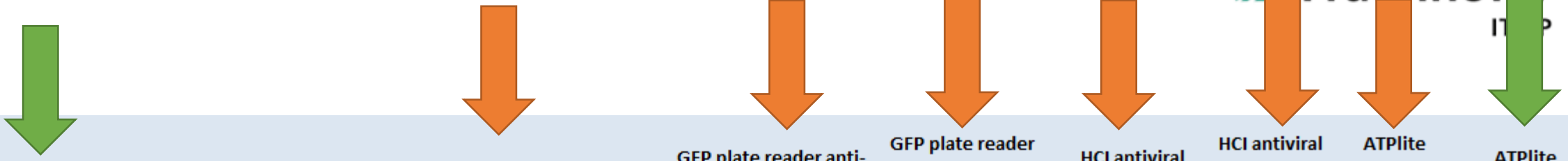
FAIR data maturity levels explained

Fraunhofer ITMP



https://fairplus.github.io/Data-Maturity/

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	SOTRASTAURIN	> 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				

https://zenodo.org/records/4774709

Examples of dataset minimal optimized

Fraunhofer ITMP

			Assay 1				
Identifier	**Synonyms**	**PubChem ID**	**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	SOTRASTAURIN	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

https://zenodo.org/records/4774709

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: VolkswagenStiftung