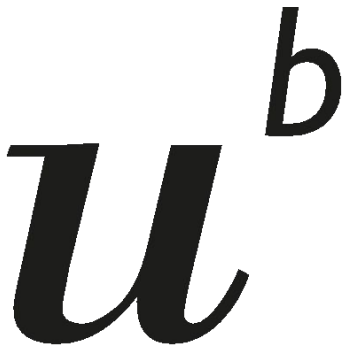


b

UNIVERSITÄT
BERN

SMART framework for FAIRness

Assisting FAIRness: SMART-metrics for FAIR research data lifecycle management

Federico Grasso Toro, Data Steward

From the point of view of ***Research Software Engineers***

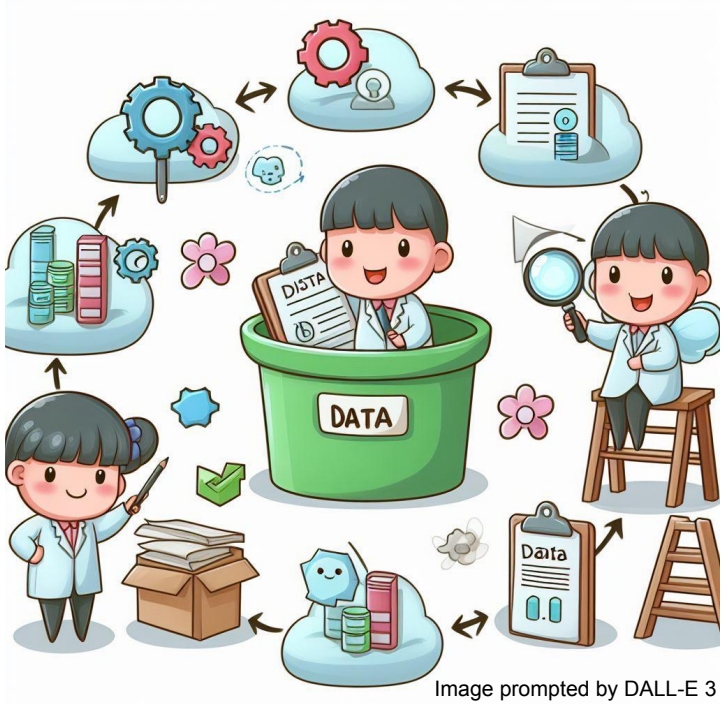


Image prompted by DALL-E 3

SMART framework for FAIRness

Content

Introduction and General Objectives

Proposed Framework – Overview

1. SMART Metrics
2. Research Data Life Cycle Management
3. FAIR Principles as Requirements
4. BDD and Gherkin Syntax

Framework Advantages

Conclusions and Further Work

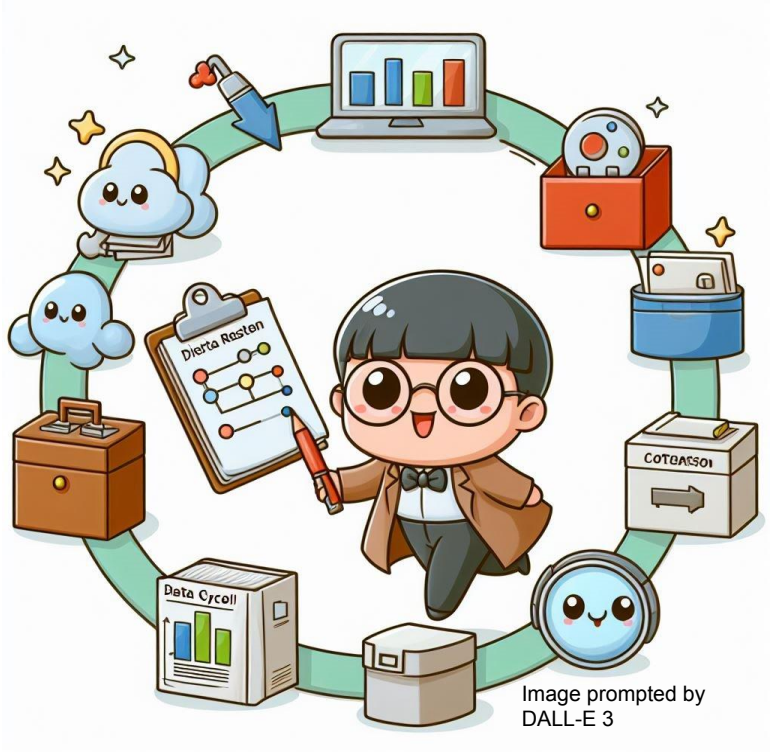


Image prompted by DALL-E 3

Data Management Plans

Introduction – A short history


Image prompted by DALL-E 3

- ➔ The concept of data management began in the 1960s, with the Association of Data Processing Service Organizations (ADAPSO) forwarding data management advice, with an emphasis on professional training and quality assurance metrics.[1]
- ➔ DMPs were originally used in 1966 to manage aeronautical and engineering projects' data collection and analysis, and expanded across engineering and scientific disciplines in the 1970s and 1980s.[2]
- ➔ Today, DMPs are widely used in research projects across various fields, including social sciences, humanities, and natural sciences.[1]

1. [DATAVERSITY] A Brief History of Data Management
2. [Burnette et al.] From Plan to Action: Successful Data Management Plan Implementation in a Multidisciplinary Project

Introduction – General Objective

FAIRness assessment by digital technologies for research data

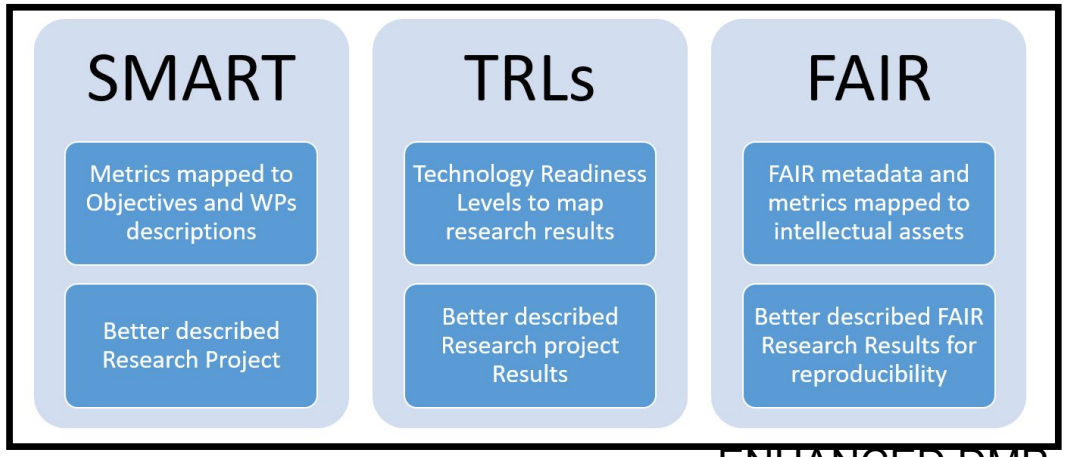
FAIR (Findable, Accessible, Interoperable, Reusable) for ensuring **reproducible science**.

➔ Challenge: Required **machine-readability** and **automating assessment process** (FAIR-metrics).

Goal:

Framework to leverage **digital technologies** to *enhance and assess* FAIRness for research results.

➔ **FAIR** is for Machines
➔ **Policies/Guides** are for humans
➔ **Frameworks** are for Human+Machine!



ENHANCED DMP

3. DMPs as Management Tool for Intellectual Assets by SMART-metrics | International Journal of Digital Curation (ijdc.net)

Proposed Framework – Overview

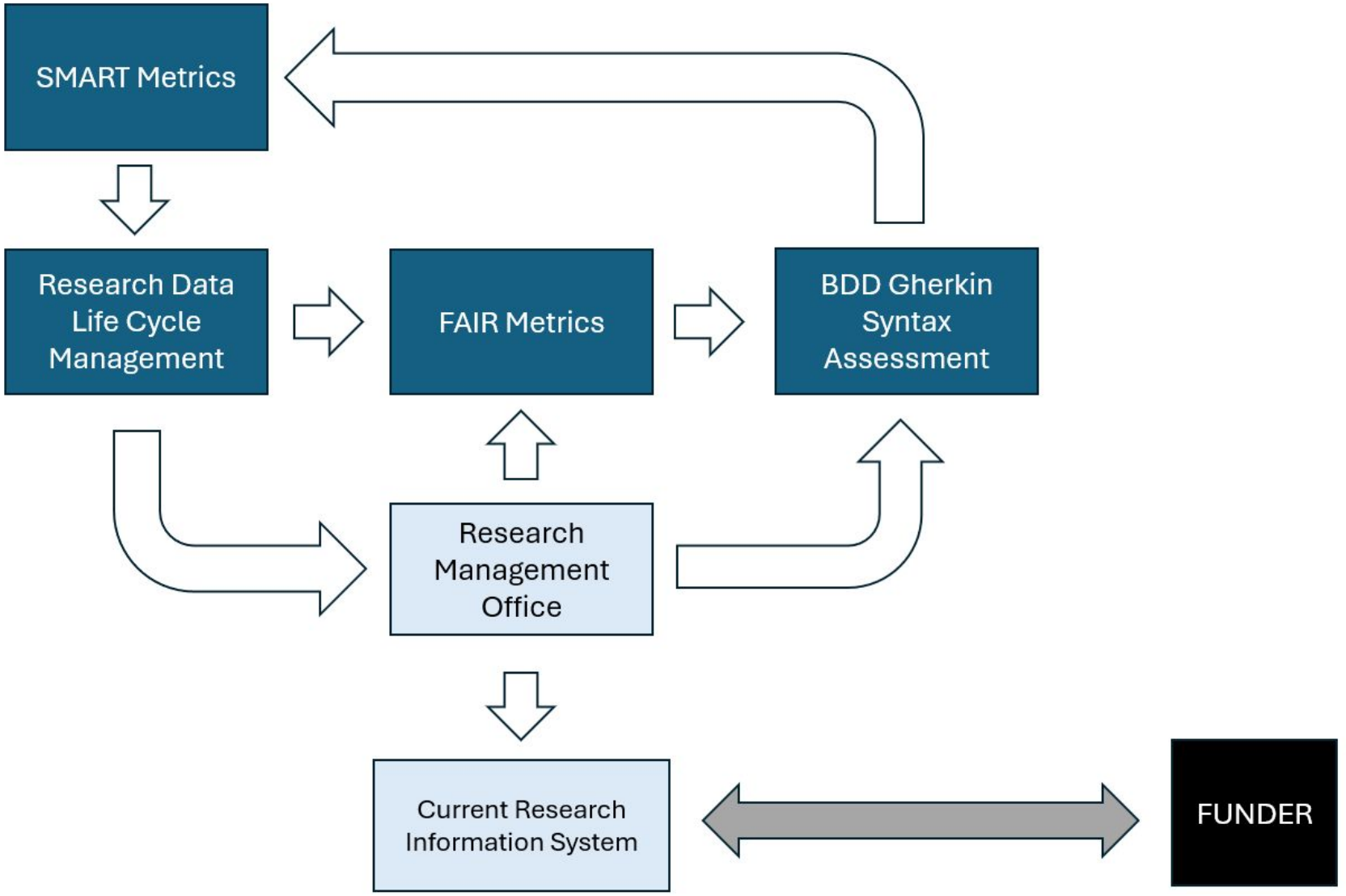
A Conceptually Sound and Technology-Agnostic Framework

Framework integrates:

1. **SMART metrics** (Specific, Measurable, Achievable, Relevant, Time-bound) for structured tracking of intellectual assets (i.e., research outcomes).
2. **Research Data Life Cycle Management (RDLCM)** to map/manage research workflow.
3. **FAIR principles as machine-requirements**, to ensure and assess compliance with established guidelines/policies.
4. **Behaviour Driven Development (BDD)** by **Gherkin syntax** for *discovery, formulation and automation* of FAIRness and any other Open Science requirements.

→ **Conceptual Strength:** Each Component supports automated FAIRness evaluation.
→ **Technology-agnostic:** Applied across various platforms and systems.

Proposed Framework – Overview



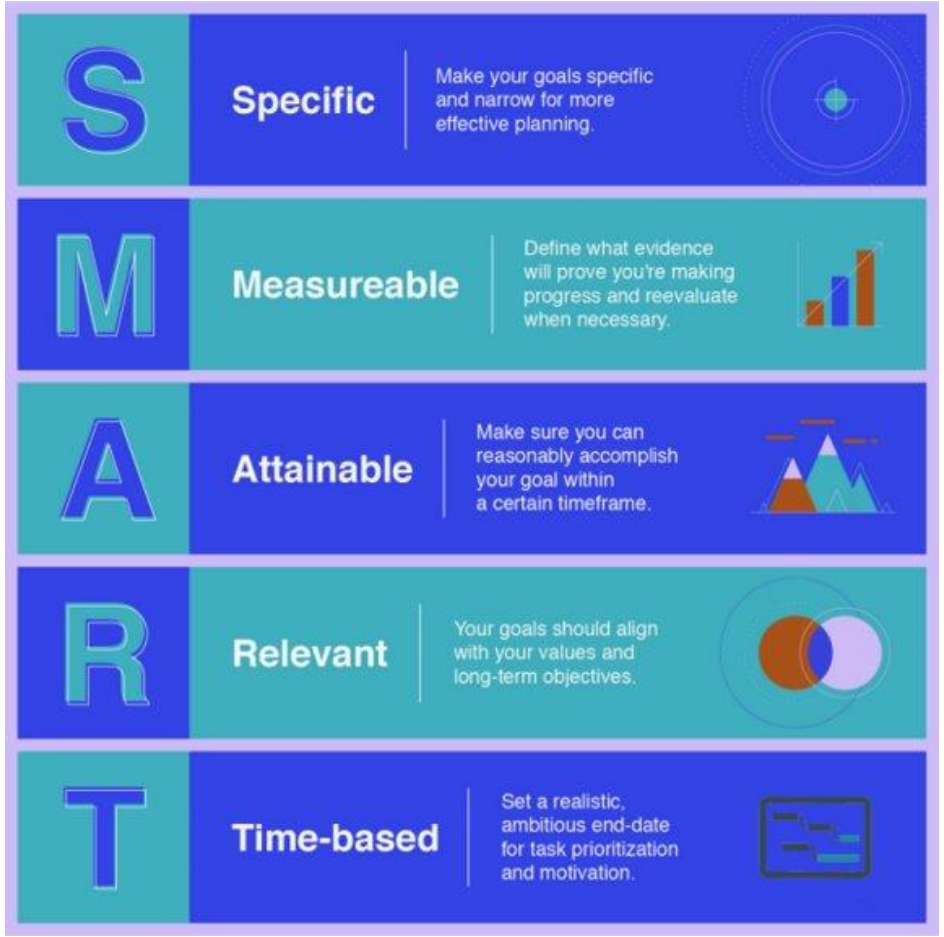
SMART Metrics

Tracking Intellectual Assets

Component 1: SMART Metrics

… criteria that help setting and measuring metrics related to *research outcomes*.

- **Specific**: Clear, defined intellectual assets.
- **Measurable**: Quantifiable, enabling assessment.
- **Achievable**: Realistic within the project's scope.
- **Relevant**: Aligned with FAIRness objectives.
- **Time-bound**: Defined **outputs** within a timeline.



4. [CFI, 2015] Corporate Finance Institute Education Inc. SMART Goals

RDLCM – Research Data Life Cycle Management

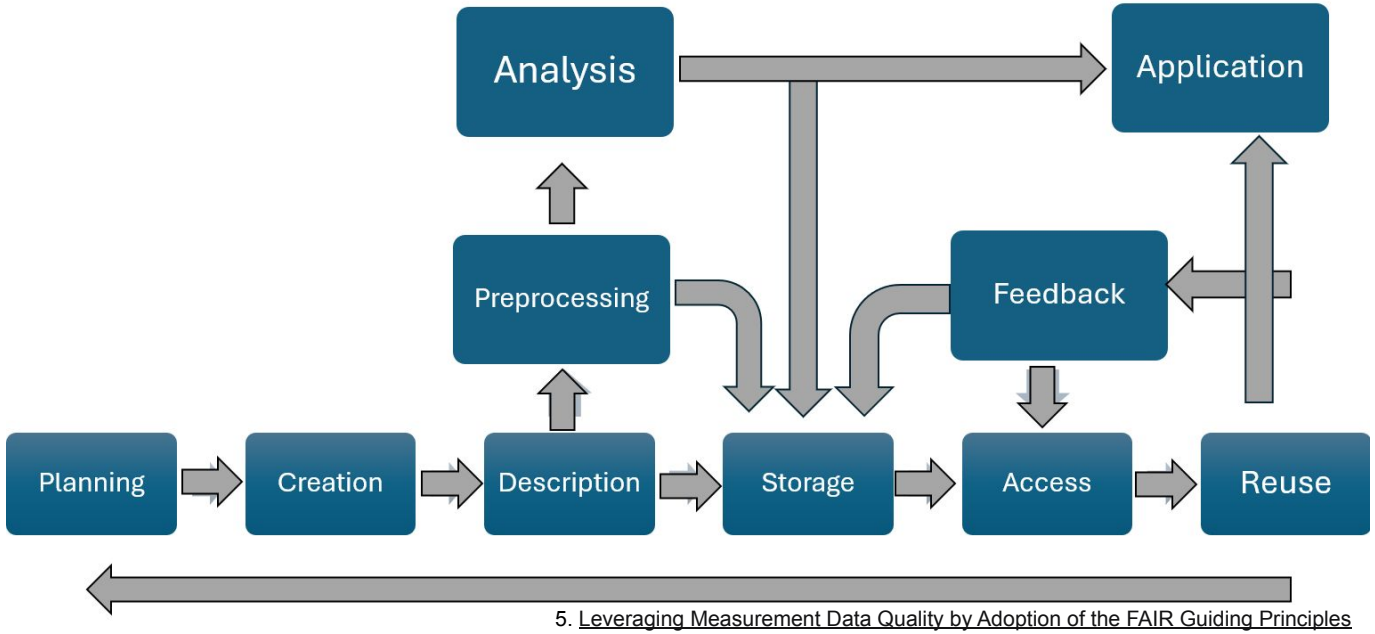
Manage research information (data + metadata)

Component 2: Research Data Life Cycle Management (RDLCM)

…the process of managing research data from creation to archiving.

General to Mapped Management

From SMART metrics RDLCM



5. Leveraging Measurement Data Quality by Adoption of the FAIR Guiding Principles
6. Leveraging measurement data quality by adoption of the FAIR guiding principles (degruyter.com)

RDLCM – Research Data Life Cycle Management

Manage research information (data + metadata)

Component 2: Research Data Life Cycle Management (RDLCM)

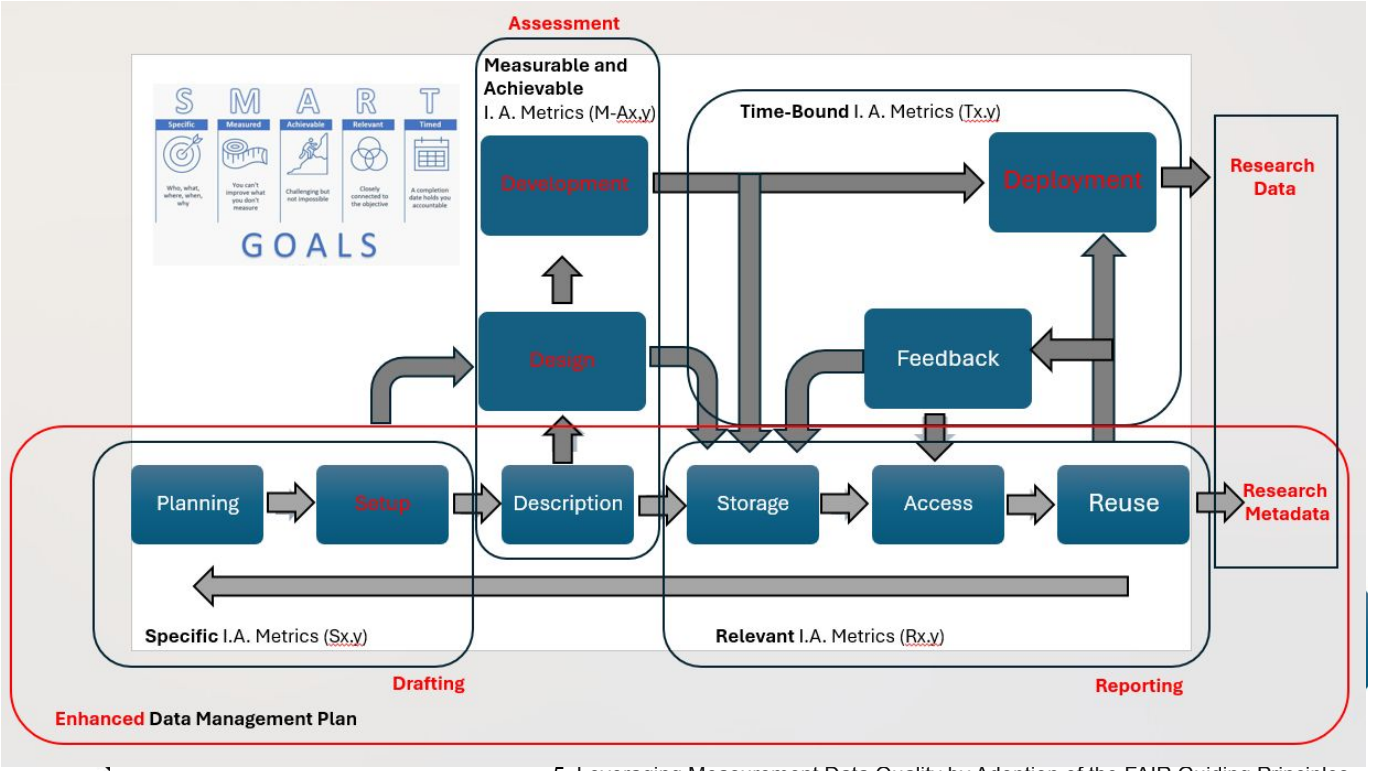
…the process of managing research data from creation to archiving.

General to Mapped Management

From SMART metrics RDLCM

Ensures Research Information are **mapped** throughout its lifecycle (creation, processing, sharing, preservation).

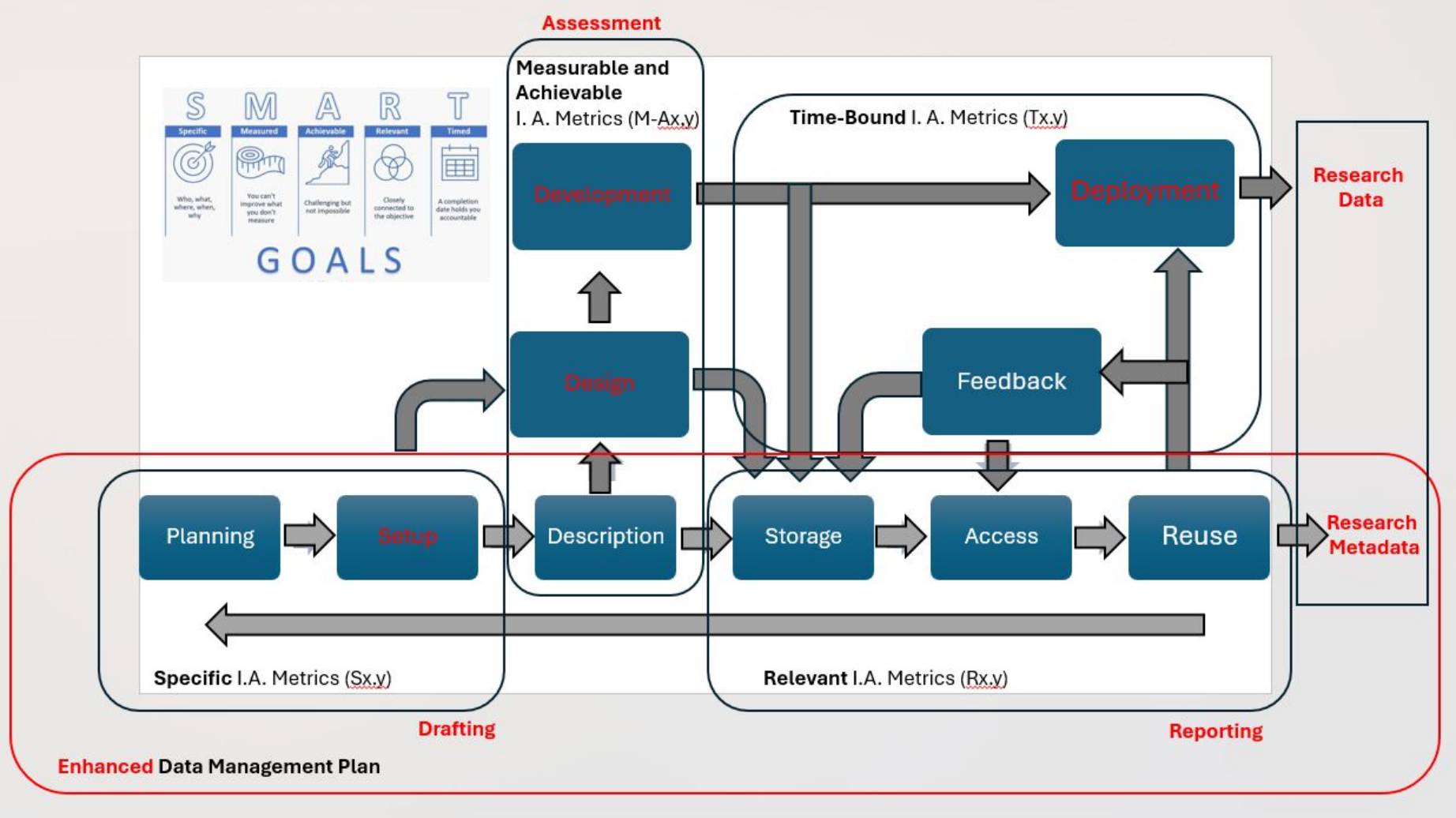
➔ Structured way to assess **if** each **phase** of the data lifecycle **aligns** with **FAIR principles** to be validated by research management office.



5. Leveraging Measurement Data Quality by Adoption of the FAIR Guiding Principles
6. Leveraging measurement data quality by adoption of the FAIR guiding principles (degruyter.com)

RDLCM – Research Data Life Cycle Management

Manage research information (data + metadata)



FAIR Principles – Subclauses as Requirements

FAIRness as references for research outcomes assesment

Component 3: FAIR Principles (as 15 requirements)

Overview:

FAIR principles addressing

findability, **accessibility**, **interoperability**,

and **reusability** of research outcomes.

➔ These 15 subclauses can be embedded as **requirements within the framework,** providing specific targets that each intellectual asset must meet.

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

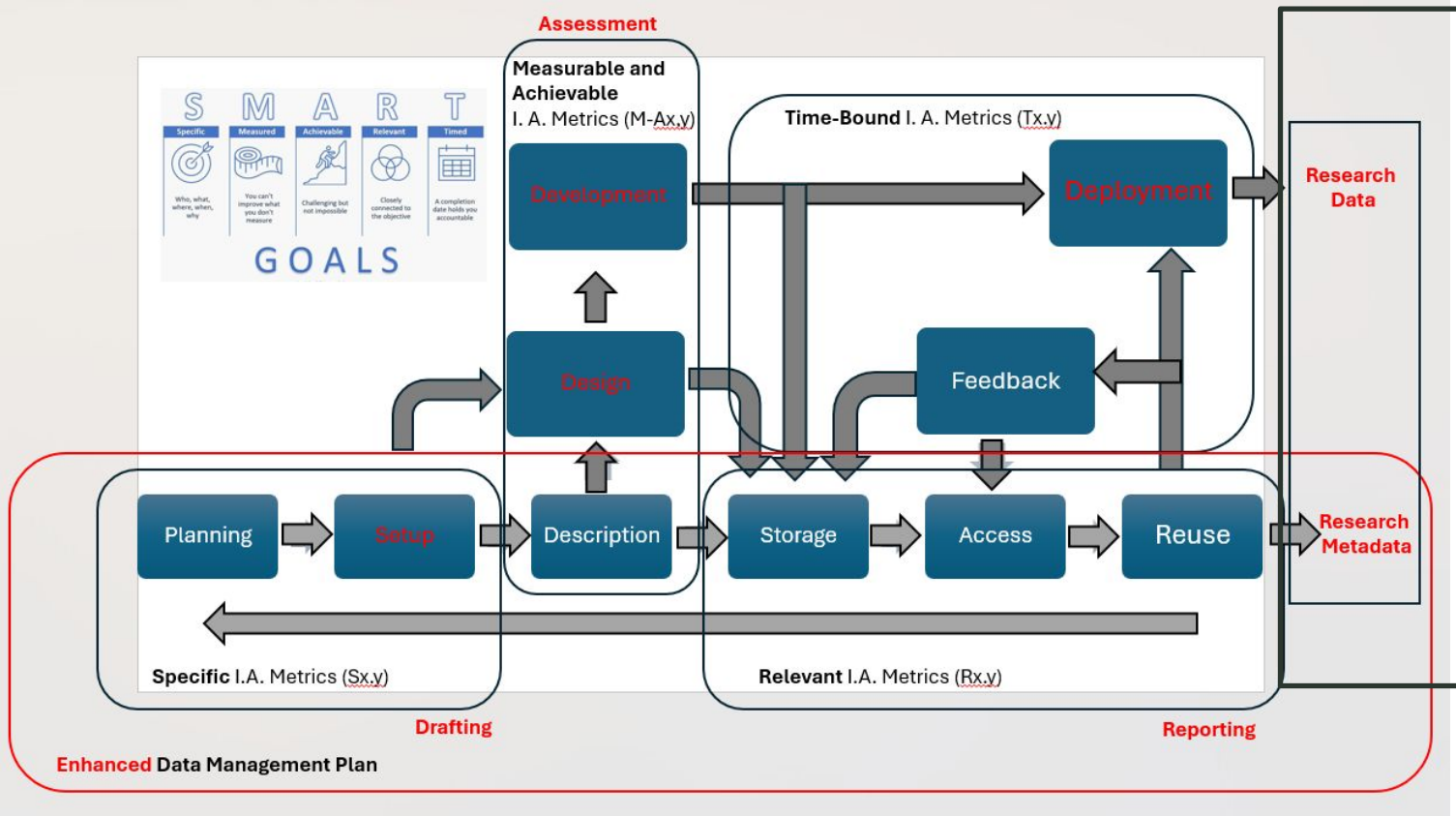
To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

7. The FAIR Guiding Principles for scientific data management and stewardship (nih.gov)

FAIR Principles – Subclauses as Requirements

FAIRness as references for research outcomes assesment



Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

7. The FAIR Guiding Principles for scientific data management and stewardship (nih.gov)

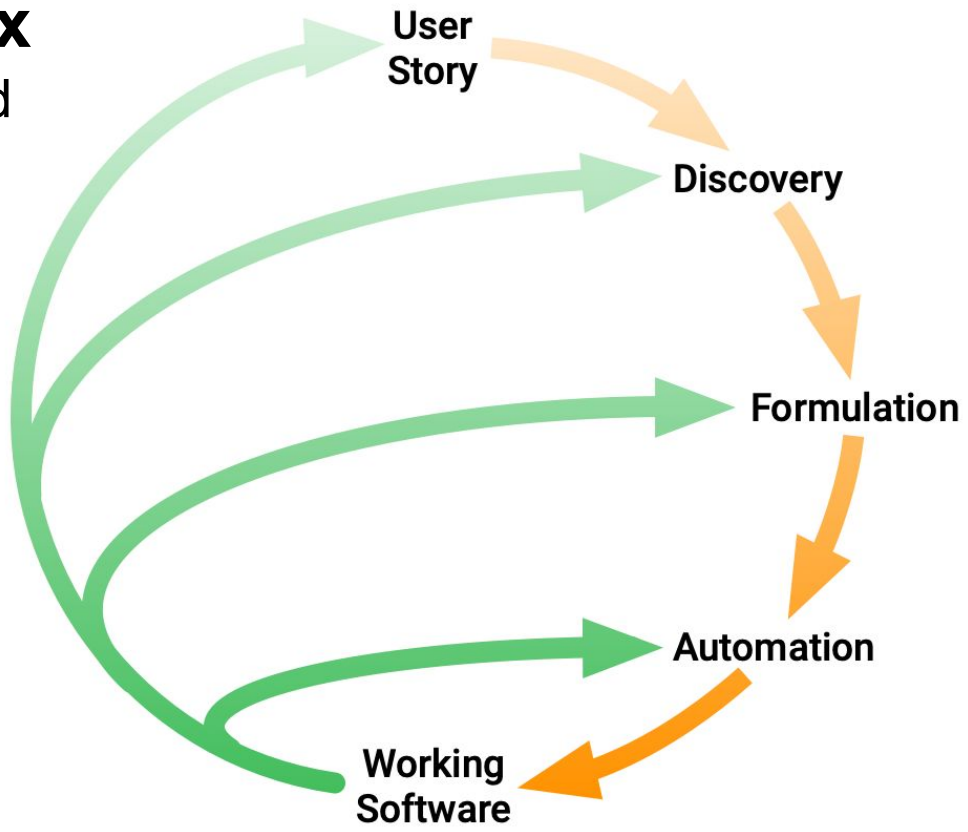
Behavior-Driven Development

Example: SMART framework - Intellectual Asset

Component 4: BDD + Gherkin Syntax

...covering the 15 FAIRness requirements, mapped into the RDLCM by human-readable language.

- Behavior-Driven Development:
  - Discovery
  - Formulation
  - Automation
- Gherkin Syntax
- Three Amigos Meetings



8. Who should formulate the scenarios? (cucumber.io)

BDD and Gherkin Syntax

Gherkin Syntax – Automating FAIR Subclauses for Each SMART Metric

Gherkin Syntax: A *human-readable* language that describes **behavioural scenarios** for systems.

It enables the automation of SMART metrics and the FAIR principles' subclauses.

- **Example Format**:
  - **Given** [initial context],
  - **When** [action is taken],
  - **Then** [expected outcome].

➔ Example for FAIR Findability (F1)
- **Given** a dataset with metadata,
- **When** the metadata is queried by a machine to collect dataset,
- **Then** the metadata (UID) must be found and solvable.

9. Gherkin Syntax - Cucumber Documentation

BDD and Gherkin Syntax

Three Amigos

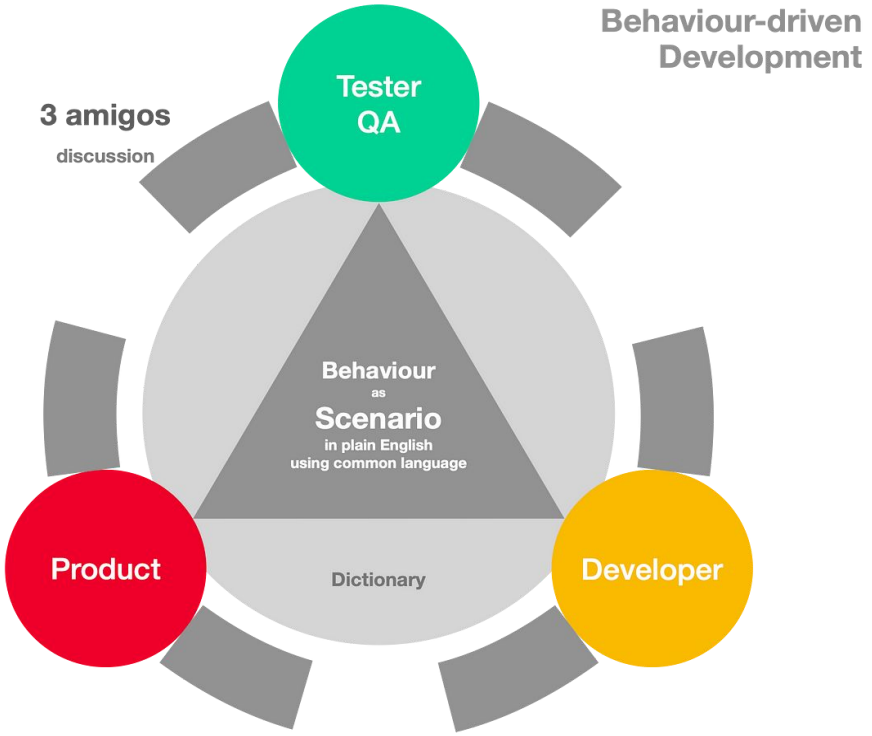
To examine work **before, during, and after** development.

Three perspectives:

- Business – What problem are we trying to solve? (**Product Owner**)
- Development – How might we build a solution to solve that problem? (**Developer**)
- Testing – What about this, what could possibly happen? (**Tester**)

These different perspectives collaborate to define:

(i) what to do, (ii) how to develop it and (iii) how to test it.

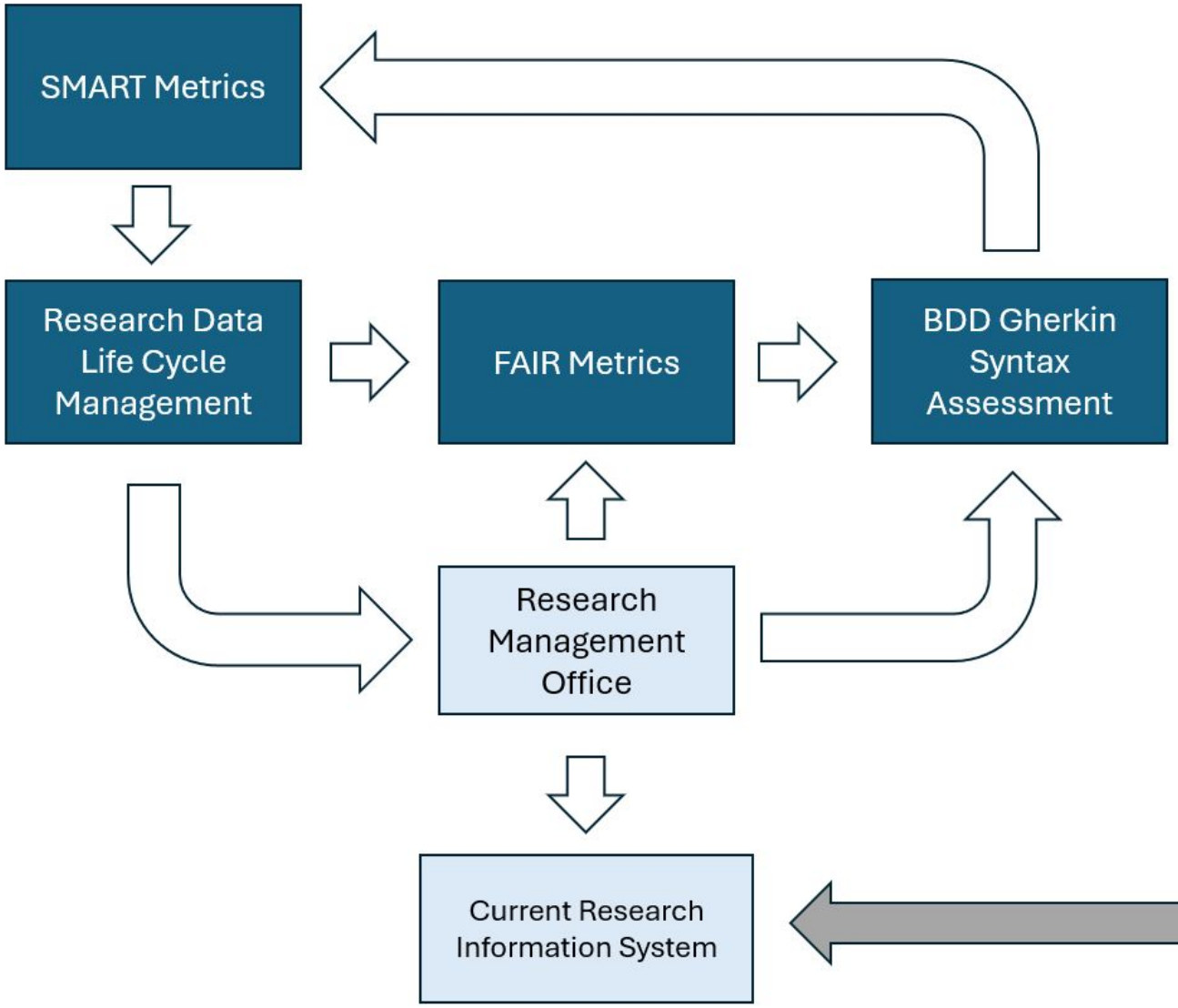


The end result?

➔ A clearer description (requirements by documentation) of an increment of work (by exemplary scenarios), leading to a shared understanding of the product.

10. What are the Three Amigos in Agile? | Agile Alliance

Proposed Framework – Overview



Key Concept:

How M2M communication plays a vital role in automating the discovery, formulation, and implementation of FAIR subclauses (FAIR Metrics).

- → BDD's machine-readable Gherkin syntax enables seamless communication between components of the framework.
- → M2M technology automates the execution of SMART Metrics across research data management platforms, reducing human intervention.
- → Traceable SMART Metrics logs are recorded by means of Data Stewardship assessment, RMO verification and CRIS validation.

Framework Advantages

Focused on Machine to Machine (maDMPs)

M2M enhances the **automation and efficiency** of the framework.

- **Scalability**: M2M allows for large-scale, **automated assessments** of FAIRness;
- **Consistency**: Ensures uniform application of SMART metrics across all intellectual asset types through RDLCM; and
- **Efficiency**: Reduces manual overhead for researchers, institutions, and funders.

Conclusions and Further Work

Partial Conclusions

➔The proposed framework offers a **holistic approach** for assessing FAIRness, integrating multiple **digital tools** and methodologies (SMART metrics, RDLCM, FAIR principles, BDD).

➔By making **data management plans "machine-readable"** and automating assessments, this approach promotes more **efficient** and **effective** research data management.

➔Highlights a shift toward leveraging **digital technologies** to streamline the entire data lifecycle (to machine-actional DMPs) from research outputs creation to assessment and evaluation.

Conclusions and Further Work

Further Work

- **Harmonization across Disciplines**: Standardizing this framework for use across different research fields.
- **Cross-Institutional Collaboration**: Encouraging adoption of this framework at national level, enabling broader **interoperability** between institutions and funders.
- **Continued Innovation**: Exploring how the framework can support not just publications and datasets, but also emerging types of **intellectual assets** (e.g., Patents, AI models, complex simulations, etc.).
- **Standardization Efforts**: Collaborating with **data governance bodies** to establish “global standards” that align with FAIR principles and other Open Science Metrics.

Contact

Dr.-Ing.

Federico Grasso Toro

Data Steward

federico.grasso@unibe.ch

+41 31 684 92 00

QR-Code generated by app

Too long, didn't read/listen?
Listen to the Human + Machine generated DS4DS-PODCAST here