SMART Metrics Based Framework For Enhanced Data Management Plans

FAIRness Assessment of Intellectual Assets

Federico Grasso Toro
University of Bern
Switzerland
federico.grasso@unibe.ch
ORCID: 0000-0002-9041-0868

Abstract **– A comprehensive framework to enhance communication between researchers, higher education institutions (HEIs) and funders is proposed for assessing the FAIRness of research results. The proposed enhanced Data Management Plans (DMPs) include: (1) Specific, Measurable, Achievable, Relevant, and Time-bound (SMART) metrics, (2) Research Data Life Cycle (RDLC) management, (3) Findable, Accessible, Interoperable, and Reusable (FAIR) principles, and (4) Behaviour Driven Development (BDD) and Gherkin syntax.**

SMART metrics provide a structured approach for planning, developing, and evaluating research outcomes, based on objective-dependent variables.

RDLC maps the assets derived from SMART metrics to assess their FAIRness, ensuring research outcomes are comprehensible to machines through enriched metadata.

The FAIR principles are operationalised in this context by enhancing the findability, accessibility, interoperability and reusability as requirements written in Gherkin syntax, following SMART metrics.

Finally, BDD facilitates their creation, updating and evaluation using asset-dependent scenarios, ensuring that all metrics are clear, actionable, and testable.

This holistic framework supports the automation, interoperability and streamlining of research data management, promoting replicability in the short term and reproducibility in the long term of research results.

Future efforts will focus on applications across different domains, i.e., automating FAIR assessments for datasets, software, and digital models, while evaluating its direct impact on improving transparency and traceability for research data management.

Keywords **– SMART, FAIR, DMP, BDD, DATA STEWARDSHIP.**

I. SMART Metrics for Open Science

In the evolving landscape of research and development, the role of Data Management Plans (DMPs) has traditionally been confined to the collection, storage, and sharing of data.

However, recent advancements and insights call for an expanded role for DMPs, known as enhanced DMPs, encompassing the management of any critical research results [1], towards the future development of machine-readable/actionable pipelines between researchers and funders. Hence, this not only broadens the scope and complexity of DMPs but also "enhances" their utility as a comprehensive research management tool.

The key elements proposed in this paper for enhancing the role of DMPs include:

- The application of SMART (Specific, Measurable, Achievable, Relevant, Time-bound) metrics [2], for defining all research outcomes that might be considered relevant intellectual assets. This is the base for the proposed framework and defines clear, measurable, and attainable project objectives. SMART metrics can facilitate a better alignment of work packages (WPs) and deliverables with the overarching goals of the project.

- The use of the research data life cycle (RDLC) management approach [3], for providing a structured method to manage any research data, from its creation until its application, ensuring that not only data but also relevant metadata remain useful and accessible, by transparent and traceable paths.

SRDD 2024: Swiss Research Data Day, Chur (Switzerland).

DOI: 10.nnnn/nnnnnnn.nnnnnnn

- The implementation of FAIR principles [4], towards defining and assessing the findability, accessibility, interoperability and reusability of the relevant research outcomes (i.e., intellectual assets).

- The development of FAIRness related requirements for each intellectual asset described by SMART metrics is drafted by means of behaviour-driven development (BDD) [5] and Gherkin Syntax Format (i.e., Given-When-Then) [6], for closing the gap between machine-to-machine types of dataflows [7] and human-centric research [8]. This helps the interpretation of objectives and deliverables between funders and researchers, while aligning with digitalisation strategies of higher education institutions (HEIs).

This framework aims to ensure that all project components are cohesively directed towards achieving their declared and desired outcomes. Hence, these enhanced DMPs will become useful tools as dynamic documents for Principal Investigators (PIs). Versioned DMPs will serve them as living records, promoting transparency and traceability of research management, while aiding the reproducibility of science. This will also enhance the overall effectiveness and accountability of funding agencies to evaluate research reports accordingly.

In the frame of Open Science, data stewardship teams will play a vital role in the enhancement of DMPs [1]. Data stewards will help researchers develop effective DMPs aligned with the drafted SMART metrics and FAIR principles, to ensure that all intellectual assets are well-defined and managed throughout the whole project, supporting the successful implementation of comprehensive research data management (RDM) and digitalisation strategies (e.g., funder's requirements institutional data governances).

Moreover, enhanced DMPs will help meeting legal and ethical standards, promoting RDM best practices, while easily fulfilling funding agency updated requirements, following novel regulations [9] [10]. By integrating SMART metrics for machine-to-machine workflow, enhanced DMPs will also become powerful tools to support the comprehensive planning, execution and assessment of research projects [1] at machine-readable level, providing valuable inputs for machine actionable DMPs (maDMPs) efforts [11].

It is important to differentiate approaches, expected impact and flexibility:

The approach from the research data alliance (RDA) [11] focuses on defining well-structured information between HEIs and funders for better reporting on research projects. While this SMART framework by enhanced DMPs aims to improve the quality of data for current research information systems (CRIS) [7] and the machine-to-machine communication between PIs, researchers and their HEIs. Hence, these SMART metric enhanced DMPs will be research data management tools for closing the loop between CRIS, HEIs and researchers, while producing machine-readable output that can be used as machine-actionable input for RDA's maDMPs approach [11].

This comprehensive framework also aims to support the successful planning, execution, and reporting of projects, ultimately contributing to the advancement of scientific knowledge and innovation by using digital technologies for advancing information science for FAIRness assessment of intellectual assets.

In conclusion, the main goal of this framework in the long term is to automate the assessment process of FAIRness by project dependent requirements.

Therefore, if FAIR principles are for machines and policies/guides are for humans, we can propose that: "Frameworks are for Human + Machine Systems".

II. RESEARCH DATA LIFE CYCLE MANAGEMENT

The importance of quality of data in addressing social/industrial challenges and ensuring sustainable research while retaining high impact research and innovation cannot be overstated [3].

Research data and any other research outcomes considered intellectual assets (e.g., software, models) serve as the backbone for numerous processes, with its reuse extending from immediate applications and related research by HEIs to long term implications for research data life cycle assessments by funders.

Despite its critical role, current research data management (RDM) faces significant challenges, often characterised by poor data quality and missing data provenance, due to inadequate documentation and/or lack of standard operating procedures (SOPs) to reinforce management practices.

This issue is exacerbated by the existence of data silos and the lack of contextual information, which collectively hinder effective data reuse [3].

To overcome these challenges, the adoption of the FAIR (Findable, Accessible, Interoperable, Reusable) guiding principles [4] is essential, especially for the future of reproducible science. These principles advocate for data to be easily locatable for both humans and machines, accessible with minimal barriers, integrable with other data and/or applications, i.e., interoperable, and well-documented to ensure long term reusability.

But implementing FAIR principles does not only ensure that data can be efficiently managed and reused. It also enhances its overall value and their corresponding impact to innovation in research, industry and society.

In response to these needs, a consolidated research data life cycle (RDLC) management framework was proposed [3], inspired by various scientific and industrial data flows, emphasising the application of FAIR principles at each stage of the data life cycle to preserve provenance metadata and to maintain high data quality and ensure reusability. An updated management framework can be seen in Fig. 1 that includes any other research outcome as intellectual assets.

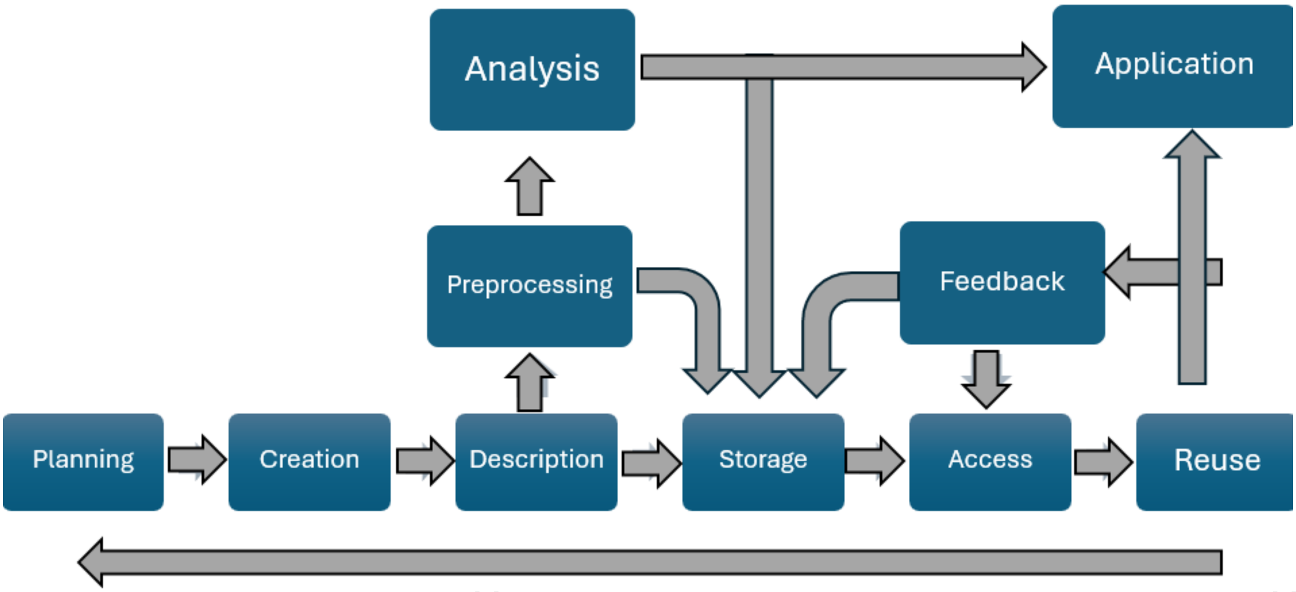

Fig. 1: RDLC management approach [3], adapted for any other research result as intellectual asset [9].

This updated RDLC management framework provides a structured approach to managing research outputs, from its creation to its final application, ensuring that all relevant intellectual assets remain as useful and accessible as possible.

By embedding the FAIR principles into the research data life cycle, one aims to elevate the quality and reusability of measurement data and any other intellectual asset. Such improvements are crucial for supporting sustainable best practices and enhancing resilience against unexpected disruptions.

The adoption of the FAIR principles and the RDLC management framework represents a significant step towards achieving high quality research outcomes, following regulatory compliance and ensuring continuous innovation and efficiency in research processes.

III. Applied SMART Metrics To RDLC

We consider the framework from [3] more than just a tool for resilient and competitive industry towards strategic autonomy in an open economy [10]. The present paper aims to further adapt this data-focused framework to scale it for any other intellectual asset [9] produced at HEIs, towards strengthening their knowledge valorisation.

Intellectual Asset are defined as: "*Any result or products generated by any Research & Innovation activities, (such as intellectual property rights, data [sets], know-how, prototypes, processes, practices, technologies, software, [models])*" [9].

Fig. 2 updates the RDLC from [3] to map SMART metrics [2] towards enhanced DMPs. This includes any relevant enriched metadata for each research outcome that can be considered an intellectual asset (e.g., data sets, software code, models).

IV. Behaviour-Driven Development Approach for SMART Metrics Drafting

The following section of the present paper explores an example for how SMART [2] metrics can be created to tackle the lacking implementation aspects for intellectual assets to be openly shared and reused, complementing the approach from [1].

The created SMART metrics can also be assessed by means of behaviour-driven development (BDD), but if and only if they were drafted following scenarios formulation [5] with structured Gherkin syntax [6].

Hence, BDD and Gherkin syntax can be combined into an effective methodology for drafting and assessing research scenarios that are aligned with expected research outcomes and milestones, ensuring that these are created as precisely as possible, while establishing interoperable means of automating and evaluating them [12].

A. *Clear Specification of Requirements*

BDD focuses on defining "requirements" in a clear and precise manner, using natural language and ensuring them to be **specific** (i.e., scenarios describe expected behaviour in detail, reducing ambiguity) and

measurable (i.e., requirements assessment can be automated by specifying expected outcomes).

B. Collaborative Approach

BDD encourages collaboration between stakeholders, including developers, testers, and domain experts. This means PIs, researchers, research software engineers (RSEs) and data stewards. Described metrics are **achievable**, **relevant** and **time-bounded** (i.e., the involved participants set realistic goals and ensure that the metrics describing the research scenarios are relevant to the project and can be practically achieved during expected timelines).

C. Automated Testing and Validation

BDD frameworks, such as Cucumber by Gherkin Syntax [6], can facilitate the automation of acceptance tests, ideal for streamlining communication between researchers, HEIs and funders.

The drafted assessment tests must consider:

- **Specific and Measurable**: Automated tests verify that the project timelines are as expected, providing measurable results to update the DMP and/or assess the specific outcomes (i.e., intellectual assets).

- **Achievable and Relevant**: Automated tests verify expected FAIRness requirements, following agreed workloads and shared objectives/deliverables descriptions, relevant to each participant and outcome.

- **Time-bound**: Automated tests can run frequently, ensuring that all relevant SMART metrics that describe the research scenario are continuously validated within specified timeframe and expected research outcome.

D. Traceability and Documentation

BDD scenarios allow DMPs to become “living documentation”, since “enhanced” DMPs provide transparency requirements (from the proposal stage of project) and traceability requirements (up to research results publishing and report stages). This also helps to ensure that all types of research results (i.e., intellectual assets) have been developed following both relevant and time-bound restrictions.

This means that the overall SMART metrics are consistently aligned with the evolving requirements and are fully tracked over time by PIs and researchers.

E. Continuous Feedback Loop

BDD supports continuous integration and delivery practices, creating a control feedback loop, where SMART metrics are constantly evaluated and refined based on the real-world performance of the research scenarios, i.e., updated DMPs become an audit trail of how intellectual assets were (1) planned, (2) researched and developed, and (3) delivered.

F. Summary of BDD for SMART metrics

BDD is a useful tool to draft and assess SMART metrics, ensuring their clarity, precision, and continuous validation for any research-related activities, while fostering a collaborative and transparent environment for increasing the effectiveness of DMPs, and supporting Open Science efforts for daily transparent, short-term replicability and long-term reproducibility.

G. Example Gherkin Syntax Format

Given [*initial context*],

When [*action is taken*],

Then [*expected outcome*].

V. PRACTICAL FRAMEWORK APPLICATION AND EXAMPLE

Here is a step-by-step method of how BDD can be used to create SMART metrics for a “machine readable” (towards “actionable”) enhanced DMP (eDMP):

A. Step-by-Step Method

The following four steps method is suggested to define the SMART metrics for enhanced DMPs.

Step 1: Discover and Draft Scenarios
Step 2: Formulate Metrics from Scenarios
Step 3: Automate Testing for Research Scenarios
Step 4: Monitor and Iterate Scenarios Descriptions

Using Gherkin Syntax Format (i.e., Given-When-Then), stakeholders can define scenarios for research data management, involving all relevant SMART metrics and following all FAIR related requirements for each specific intellectual asset.

This 4-steps method provides a structured approach to implementing Behaviour-Driven Development (BDD) for robust and reliable software:

STEP 1: Discover and Draft Scenarios

Stakeholders (PIs, researchers, research software engineers and data stewards) collaborate to identify key scientist behaviours and translate them into concrete, detailed scenarios written in a clear and unambiguous language (by following Gherkin Syntax Format).

This ensures all stakeholders involved in the drafting of the enhanced DMPs understands desired behaviour (e.g., research inputs/outputs and intellectual assets).

STEP 2: Formulate Metrics from Scenarios

Stakeholders define measurable outcomes for each scenario to objectively assess their success. The resulting SMART metrics can include performance benchmarks, research milestones and/or specific functional research outcomes as intellectual assets.

This step connects scenarios to quantifiable goals.

STEP 3: Automate Testing for Research Scenarios

Stakeholders transform the drafted scenarios into automated tests, following SMART and FAIR metrics.

This allows for continuous validation of the RDLC against the defined objectives, as well as for rapid feedback and early detection of deviations.

STEP 4: Monitor and Iterate Scenario Descriptions

Stakeholders continuously track the effectiveness of scenarios through the gathered metrics, refining and updating research scenarios based on changing requirements and performance analysis.

This ensures that the research scenarios and expected outcomes remain relevant and valuable.

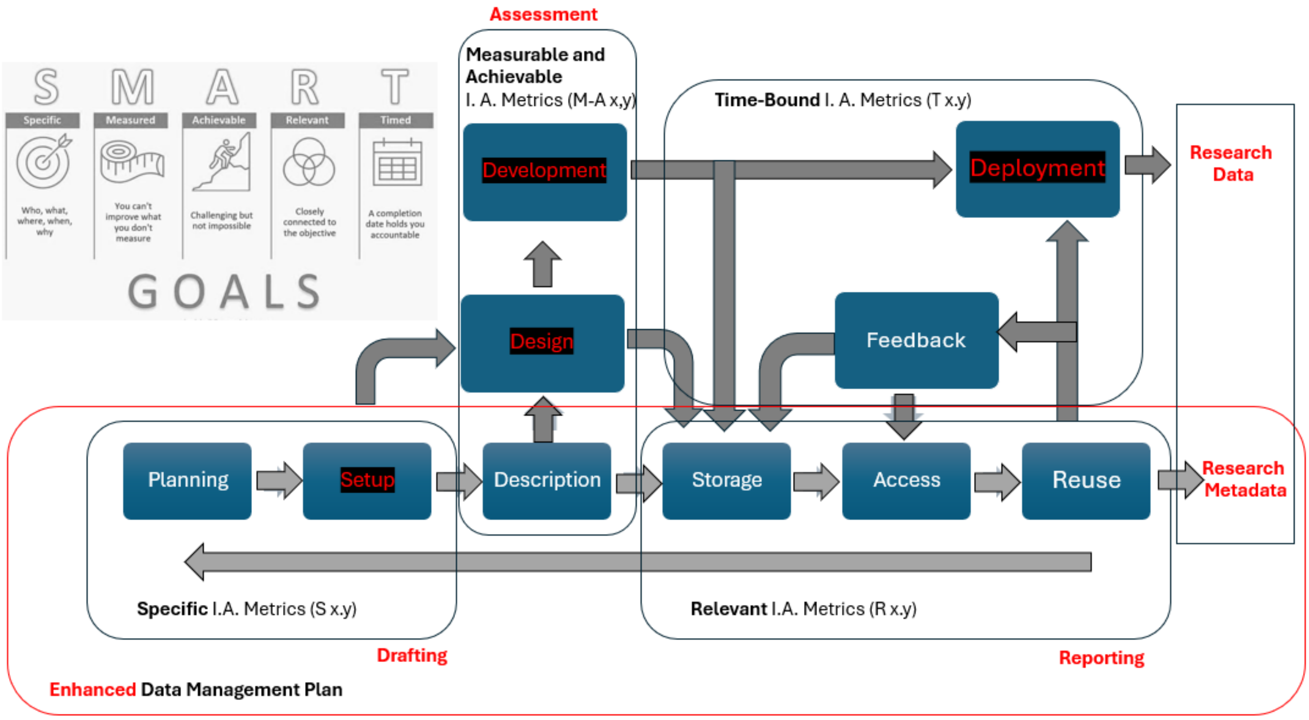

Fig. 2: Mapped SMART metrics [2] onto adapted RDLC [3] for providing enriched metadata for FAIRness [4].

B. *FAIRness related requirements examples*

This section includes FAIR requirements that can be easily adopted as FAIR metrics to assess the FAIRness of an intellectual asset (e.g., a dataset):

1. FAIR Findability (F1)

- Given: a dataset with metadata,

- When: the metadata is queried by a machine to collect the dataset,

- Then: the required metadata must be found and resolvable, e.g., DOI (Digital Object Identifier).

2. FAIR Accessibility (A1)

- Given: a dataset with proper access controls,

- When: a user/machine queries the dataset via its DOI (Digital Object Identifier),

- Then: the dataset must be accessible, and the necessary permissions (e.g., authentication) must be in place to retrieve it.

3. FAIR Accessibility (A2)

- Given: a dataset with an established access policy,

- When: a user/machine requests access,

- Then: the dataset's access policy must be clear and machine-readable (e.g., through metadata and/or rights statements) before access is granted.

4. FAIR Interoperability (I1)

- Given: a dataset with structured metadata (e.g., in a standard format like JSON-LD or RDF),

- When: the metadata is queried by a machine using an API,

- Then: the machine must be able to interpret and extract relevant information from the metadata, regardless of the platform or software used.

5. FAIR Interoperability (I2)

- **Given:** a dataset using standard vocabularies for its metadata (e.g., DDI for social sciences or OBO for biology),

- **When:** the dataset is accessed via an interoperable platform and/or API request,

- **Then:** the dataset must be correctly interpreted using the relevant standard vocabularies, ensuring consistency across systems.

6. FAIR Reusability (R1)

- **Given:** a dataset with comprehensive enriched metadata describing the context and methodology of the data collection procedures,

- **When:** a user/machine queries the dataset via an interoperable platform and/or API request,

- **Then:** the dataset must be reusable, with sufficient information - data plus enriched metadata - for users to understand how to interpret and use the data.

7. FAIR Reusability (R2)

- **Given:** a dataset with a defined license for reuse (e.g., CC-BY or similar),

- **When:** the dataset is accessed by user/machine,

- **Then:** the dataset must be available for reuse under the specified terms and conditions, and those terms must be also machine-readable.

VI. Implementation By Data Stewardship

The presented framework, encompassing enhanced DMPs through SMART metrics, applied FAIRness for data and/or any other intellectual asset life cycle(s), as well as BDD-based research scenarios, marks a significant advancement in RDM practices [1].

Its implementation will rely mostly on data stewards at HEIs level, since they will play a dynamic role within this framework, becoming key facilitators in (1) requirements gathering, (2) collaboration between PIs, researchers and research software engineers (RSEs), and (3) SMART/FAIR metrics validation.

A simplified list of activities that data stewards will have to develop, as part of their collaboration with relevant stakeholders, include:

1. Define precise FAIR requirements for specific intellectual assets (e.g., datasets or software).
2. Bridge communication gaps between PIs, researchers, RSEs and CRIS/funders.
3. Ensure SMART metrics by feasibility studies.
4. Automation/monitoring of enhanced DMPs.
5. Provide transparent feedback for RDLC.
6. Draft and update documentation to ensure traceability and transparency of research.
7. Promote standardisation between funders and HEIs by providing training/tutorials to aid enhancing DMPs and other RDM practices.

VII. Future Applications And Use Cases

The present approach enhances the precision and evaluability of SMART/FAIR metrics while fostering transparency, reproducibility, and collaboration between researchers, HEIs, and funding agencies.

The adapted RDLC framework, alongside FAIR-aware measures, aims to facilitate high-quality data reuse, ensuring long-term value and compliance with current and future regulatory requirements, such as the Digital Product Passport [13].

By integrating FAIR principles into intellectual assets management their visibility, usability, and longevity will boost, promoting a collaborative and reproducible scientific environment.

The machine-actionable DMPs proposed by RDA [11] will be crucial for the future of science. These tools must be designed to enhance communication between researchers and funders, improving the tracking and management of research processes. Hence, their development should also focus on standardisation and integration with existing internal systems (e.g., CRIS), where pilot projects with funding agencies can establish best practices based on relevant HEIs' feedback. The present framework can be applied to various use cases in several domains:

- **Clinical trials:** Tracking sensitive data and ensuring compliance with GDPR [14] while properly sharing methods for replicability.
- **Environmental science:** Monitoring the long-term ecological changes [15] while facilitating their data, models and ontologies sharing.
- **Digital humanities:** Managing complex datasets of textual and multimedia materials for analysis and preservation [16] towards transparent and traceable approaches comparisons.

Future efforts should also aim to enhance DMPs by broadening their scope to include a wider range of intellectual assets.

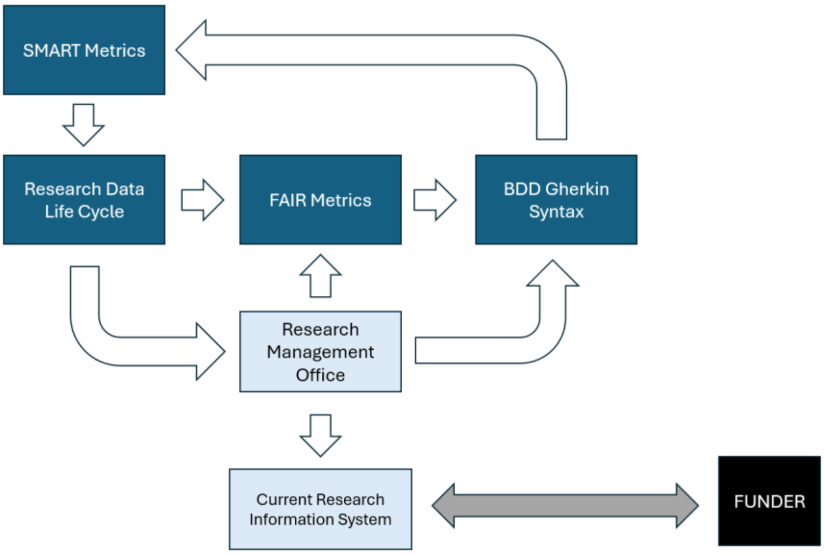

Fig. 3: Proposed framework to improve communication between researchers, institutions and funders.

VIII. CONCLUSIONS

This paper presents an overall framework, encompassing enhanced DMPs by (1) SMART metrics, (2) applied FAIR principles for intellectual assets life cycle management and (3) the application of BDD for the description of research scenarios, marks significant advancements in the RDM approach from [1].

Fig. 3 depicts how each element of the framework aids communication at machine-to-machine level between researchers, HEIs and funders. At researchers' level: (1) SMART Metrics, (2) RDLC, (3) FAIR Metrics, and (4) BDD Gherkin Syntax, while at HEIs' level: (5) Research Management Office, as well as (6) Current Research Information System (CRIS).

By integrating all these components, the framework promotes a more structured, transparent, and traceable collaborative approach to research data management. This not only enhances the quality and reusability of research outputs such as data and software code, but also facilitates better communication and coordination between several types of stakeholders.

The framework's emphasis on machine-readability further contributes to the automation and streamlining of research data management processes, ultimately supporting replicable research in the short term and reproducible outcomes in the long term.

While the framework holds significant promise, its successful implementation will rely mostly on the active participation and collaboration between researchers, data stewards, and funding agencies.

Further future efforts should focus on developing comprehensive training materials and support resources to facilitate the greater adoption of the framework across diverse research domains, e.g., [17].

Additionally, field-oriented research is needed to evaluate the long-term impact of the framework on research practices and outcomes. Since, despite the challenges that lie ahead, the proposed framework might represent a critical step toward realising the full potential of Open Science initiatives. By fostering a more coordinated and transparent approach to research, this framework can also contribute to the digital trust for advancement on scientific knowledge and innovation.

REFERENCES

[1] Grasso Toro, F. (2024). DMPs as Management Tool for Intellectual Assets by SMART-metrics. IJDC, 18. https://doi.org/10.2218/ijdc.v18i1.919

[2] Bjerke, M. B., & Renger, R. (2017). Being smart about writing SMART objectives. Evaluation and program planning, 61, 125–127. https://doi.org/10.1016/j.evalprogplan.2016.12.009

[3] Schmitt, R., Bodenbenner, M., Hamann, T., Sanders, M., Moser, M. & Abdelrazeq, A. (2024). Leveraging measurement data quality by adoption of the FAIR guiding principles. tm - Technisches Messen, 91(9), 441-453. https://doi.org/10.1515/teme-2024-0040

[4] Wilkinson, M. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 160018. https://doi.org/10.1038/sdata.2016.18

[5] Rose, S. (2019). BDD Books 2 - Formulation. Extract: BDD: Who should formulate the scenarios? https://cucumber.io/blog/bdd/who-should-formulate-the-scenarios/

[6] SMART BEAR (2019). Gherkin Syntax - Documentation. https://cucumber.io/docs/gherkin/

[7] Farinelli, C., & Zigoni, A. (2022). Extending the value of a CRIS with research data management.

Procedia Computer Science, 211, 187–195. https://doi.org/10.1016/j.procs.2022.10.190

[8] Townsend, M. and Romme, G. (2024). The Emerging Concept of the Human-Centered Organization: A Review and Synthesis of the Literature. Humanistic Management Journal. 9. https://doi.org/10.1007/s41463-024-00168-w.

[9] Commission Recommendation (EU) 2023/499: Code of Practice on the management of intellectual assets for knowledge valorisation in the European Research Area http://data.europa.eu/eli/reco/2023/499/oj

[10] Commission Recommendation (EU) 2024/736: Code of Practice on citizen engagement for knowledge valorisation http://data.europa.eu/eli/reco/2024/736/oj

[11] Miksa, T. et al. (2020) RDA DMP Common Standard for Machine-actionable Data Management Plans. http://doi.org/10.15497/rda00039

[12] SMART BEAR (2021). Cucumber Studio Documentation. Available online.

[13] Henrique, P. (2022). Brand Book CIRPASS Project. Available online.

[14] Regulation (EU) 2016/679: http://data.europa.eu/eli/reg/2016/679/oj

[15] Haase, P., et al. (2018). The next generation of site-based long-term ecological monitoring: Linking essential biodiversity variables and ecosystem integrity. Science of The Total Environment, 613-614, 1376–1384. https://doi.org/10.1016/j.scitotenv.2017.08.111

[16] Hassani, Hossein et al. (2020). Text Mining in Big Data Analytics. Big Data and Cognitive Computing. 4. 1. https://doi.org/10.3390/bdcc4010001.

[17] Barbot, L., et al. (2019). SSHOC D7.1 system specification - SSH open marketplace (1.0). Zenodo. https://doi.org/10.5281/zenodo.3547649