FAIR in action - a flexible framework to guide FAIRification

Danielle Welter[1] (0000-0003-1058-2668), Nick Juty[2] (0000-0002-2036-8350), Philippe Rocca-Serra[3] (0000-0001-9853-5668), Fuqi Xu[4] (0000-0002-5923-3859), David Henderson[5] (0000-0002-6433-200X), Wei Gu[1] (0000-0003-3951-6680), Jolanda Strubel[6,] (0000-0001-6675-4639), Robert T. Giessmann[5,7] (0000-0002-0254-1500), Ibrahim Emam[8] (0000-0002-7561-2787), Yojana Gadiya[9] (0000-0002-7683-0452), Tooba Abbassi-Daloii[10] (0000-0002-4904-3269), Ebtisam Alharbi[11] (0000-0002-3887-3857), Alasdair J G Gray[12] (0000-0002-5711-4872), Melanie Courtot[4,13] (0000-0002-9551-6370), Philip Gribbon[9] (0000-0001-7655-2459), Vassilios Ioannidis[14] (0000-0002-4209-2578), Dorothy S. Reilly[15] (0000-0002-6677-3132), Nick Lynch[16] (0000-0002-8997-5298), Jan-Willem Boiten[17] (0000-0003-0327-638X), Venkata Satagopam[1] (0000-0002-6532-5880), Carole Goble[2] (0000-0003-1219-2137), Susanna-Assunta Sansone[3] (0000-0001-5306-5690), Tony Burdett[4] (0000-0002-2513-5396).*

*corresponding author: Tony Burdett (tburdett@ebi.ac.uk)

Nr	Author affiliations
1	Luxembourg Centre for Systems Biomedicine, ELIXIR Luxembourg, University of Luxembourg, L-4367 Belval, Luxembourg
2	University of Manchester, Department of Computer Science, The University of Manchester, Manchester, M13 9PL, UK
3	Oxford e-Research Centre, Department of Engineering Science, University of Oxford, 7 Keble Road, OX13QG, Oxford, UK
4	European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton CB10 1SD, UK
5	Bayer AG, Business Development & Licensing & OI, Muellerstrasse 178, 13353 Berlin, Germany
6	The Hyve BV, Arthur van Schendelstraat 650, 3511 MJ Utrecht, The Netherlands
7	Institute for Globally Distributed Open Research and Education (IGDORE)
8	Data Science Institute, Imperial College, London
9	Fraunhofer Institute for Translational Medicine and Pharmacology (ITMP) and Fraunhofer Cluster of Excellence for Immune Mediated Diseases (CIMD), Schnackenburgallee 114, 22525 Hamburg, and Theodor Stern Kai 7, 60590 Frankfurt, Germany

10	Department of Bioinformatics (BiGCaT), NUTRIM, FHML, Maastricht University, Maastricht, The Netherlands.
11	College of Computer and Information Systems, Umm Al-Qura University, Mecca, SA
12	Department of Computer Science, Heriot-Watt University, Edinburgh, EH14 4AS, Scotland, UK
13	Ontario Institute for Cancer Research MaRS Centre, 661 University Avenue, Suite 510, Toronto, Ontario, Canada M5G 0A3
14	Vital-IT Group, SIB Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
15	Novartis Institutes for BioMedical Research, Novartis Pharma AG, Basel, Switzerland
16	OpenPhacts Foundation, Cambridge, UK
17	Foundation Lygature, Utrecht, Netherlands

Abstract

The COVID-19 pandemic has highlighted the need for FAIR (Findable, Accessible, Interoperable, and Reusable) data more than any other scientific challenge to date. We developed a flexible, multi-level, domain-agnostic FAIRification framework, providing practical guidance to improve the FAIRness for both existing and future clinical and molecular datasets. We validated the framework in collaboration with a wide range of public-private partnership projects, demonstrating and implementing improvements across all aspects of FAIR, using a variety of datasets, to demonstrate the reproducibility and wide-ranging applicability of this framework for intra-project FAIRification.

Introduction

The past two years have exposed how critical interoperability of data and systems are to society in times of crisis. The deadly COVID-19 pandemic has made people acutely aware of the weak points that have been known to data management experts for a long time: service incompatibilities, data access restrictions, unavailability of data, missing data, and incomplete, ambiguous or absent metadata. These deficiencies have plagued the scientific endeavour, in both academia and industry, and have hampered the management of the COVID-19 crisis in the early stages, from lack of transparency on data provenance (https://www.bbc.com/news/technology-54423988) to difficulty of data sharing, both due to the sensitive nature of personal health data and interoperability issues between different data sources[1,2]. These issues brought to the forefront the call to arms made in the 2016 publication about the “FAIR (Findable, Accessible, Interoperable, Reusable) Data Principles”, in which Wilkinson and colleagues[3] highlighted with an elegant acronym how life sciences data and services should be improved in order to build an infrastructure for the 21st century.

So successful was the initiative, that it was incorporated into the G20 Leaders' Communiqué from the Hangzhou Summit (https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_16_2967) and made a priority by many research funding organisations, including the Horizon 2020 programme of the European Commission (https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf). Despite uptake at the policy level and the known benefits of the FAIR principles, detailed technical guidance towards their implementation is still lacking. Feedback from the data management frontlines indicates that there is a significant demand for hands-on, practical advice on how to translate general and high-level FAIR principles into actionable, "tried and tested" processes.

This manuscript describes a "FAIRification framework" designed to address this demand by supporting organisations and projects undertaking a FAIR transformation. Specifically, we describe a reproducible and sustainable process that can be used to improve the adoption of the FAIR principles by optimising the use of available resources and expanding organisational FAIR data management capabilities. This is achieved through focused prioritisation of needs, based on a thorough analysis of the unique and specific FAIR challenges of each specific project. This framework is one of the outcomes of the FAIRplus consortium (https://fairplus-project.eu), an international project with partners from academia and major pharmaceutical companies, funded by the Innovative Medicines Initiative (IMI, https://www.imi.europa.eu), the largest private-public partnership program funding health research and innovation.

Results

Our FAIRification framework (https://w3id.org/faircookbook/FCB079) consists of three distinct components, shown in Figure 1: a reusable FAIRification Process, which outlines the main phases of a FAIRification activity; a FAIRification Template, which breaks down key elements of the process into a series of steps to follow when undertaking a FAIR transformation; and a FAIRification Workplan layout, which provides a structure for organising FAIR implementation work tailored to the needs of a specific project. Our FAIRification framework was developed in collaboration with over 17 IMI data-producing research projects[4] (full list in Supplementary Table 1). Throughout these numerous collaborations, we applied this framework to clinical interventional study datasets, data generated in the laboratory to elucidate molecular interactions, as well as real-world and clinical observational data. However, the framework is generalizable to any dataset, as well as other disciplines beyond the life sciences. Its power lies in providing a process that is reproducible over and over, making it a sustainable solution for any organisation looking to improve its FAIR data processes. The framework is also agnostic of any specific implementation solutions or methodologies, allowing users to leverage the resources already at their disposal rather than being forced to invest in solutions that may not be right for them.

Importantly, whilst we describe a reproducible framework for undertaking a FAIR transformation, we do not seek to demonstrate that our framework provides the best possible approach to FAIRification in all contexts: comparison of FAIRification results across projects and domains is a complex activity outside the scope of this work. Furthermore, although there exist a small number of FAIRification processes, such as the GO FAIR Three-point FAIRification Framework (https://www.go-fair.org/how-to-go-fair/), these tend to

focus on adoption of a specific technical stack (FAIR Data Points in the case of the GO FAIR framework), rather than providing a comparable guided transformation, and as such results cannot be directly evaluated.

The four phases of the FAIRification Process

The FAIRification Process, outlined in Figure 2 and Supplementary Figure 2, describes the general steps that should be followed when engaging in any FAIRification activity. It consists of four distinct phases: a goal definition phase, an initial project examination phase, an iterative cyclical FAIRification phase and a post-FAIRification review.

To validate the reproducibility of the process we developed, we evaluated FAIRness improvements for the datasets from the participating IMI projects by comparing dataset maturity[5] (see Methods and https://fairplus.github.io/Data-Maturity/) before and after FAIRification. A summary of the evaluation, shown in Figure 3, clearly indicates that FAIRness and maturity improved for all projects that were subjected to our methodology. It is however important to note that maturity levels should not be used to compare across different projects as results depend on a number of factors that can be highly specific to individual projects and datasets, with some indicators not being applicable to all projects. The indicators should serve only to highlight areas for improvement prior to FAIRification and provide an illustration of the scale and impact of the improvements once implemented, for a single project.

Phase 1: Set realistic and practical goals

Before any FAIRification work is undertaken, it is necessary to determine the desired usability of the data that is not achievable in its current state. From this, one or more clear and precise FAIRification goals can be determined. Based on our experience with the IMI projects, aiming to improve every aspect of FAIR is neither useful nor desirable. All FAIRification efforts come at a cost but may not yield equal levels of benefit[6]. We therefore recommend defining an acceptable “FAIR enough” end state for a dataset whereby key required capabilities are obtained while smaller, less useful enhancements are disregarded. Our experience also suggests that good FAIRification goals should have a defined scope and clearly state how the work will improve scientific value, as well as be specific and actionable.

As visualised in Figure 3, the CARE (http://www.imi.europa.eu/projects-results/project-factsheets/care) dataset increased in maturity from level 0 or “single-use data level”, to level 3, or “community level”, thanks to a clear objective of improving access to data and its overall discoverability: “*To make the project’s bioactivity data comply with community standards and publicly available so that other researchers can easily reuse the data without repeating the compound identification and testing work.*”. This goal clearly states an aim (compliance with community standards and public availability of data), a scope (the bioactivity data), and the expected scientific value (easily reuse the data without repeating the compound identification and testing work). The CARE aims were implemented in targeted interventions, such as adding an explicit license to the dataset and submitting data to ChEMBL[7], an international chemical and bioassay repository that generates FAIR-compliant (i.e. globally unique, persistent and resolvable) identifiers and indexed searchable metadata.

We recommend avoiding the word “FAIR” and its derivatives in goals entirely as it is too general to impart clear meaning, e.g. “*FAIRify data resource for public release on project platforms*”. Unlike CARE’s goal, this one does not specify the aim or scope of the work such as compliance with a community standard, mapping to controlled vocabularies or assignment of unique, persistent identifiers. The scientific value is purely implicit - submission to public platforms would likely increase findability - but is not explicitly stated. Finally, FAIRification goals should not factor in implementation details such as how the goal will be accomplished or implemented in terms of resource availability and technologies. These will be considered in the next phase.

Phase 2: Carefully examine data, capability and resource requirements

Alongside and following the goal-setting step, we identified the need for a project examination process. From early pilot projects such as ND4bb (http://www.imi.europa.eu/projects-results/project-factsheets/nd4bb) and Onco Track (http://www.imi.europa.eu/projects-results/project-factsheets/onco-track), we learned that FAIRification was difficult to accomplish successfully if project capabilities and resources were not fully understood from the outset. For example, goals relating to data hosting improvements cannot be fulfilled if data is not available or accessible, or if the project partners have not reached an agreement on the appropriate licensing and data use conditions. Similarly, goals targeting the annotation of data with open terminologies are only implementable if the data is sufficiently well understood to identify suitable controlled vocabularies and ontologies, and if expertise is available to perform the annotation to a sufficiently high standard.

Tasks related to project examination can be broken down into two distinct categories. Requirements related to the characterization of data such as data types, identifiers, metadata and data standards are categorized as “data requirements” tasks. These tasks are expected to have varying levels of complexity depending on the maturity level targeted for the dataset. Identifying the characteristics that a FAIR dataset should exhibit based on the previously defined FAIRification goal, such as conforming to a specific community standard, has been explicitly added as part of the project examination phase of the FAIRification process.

The tasks in the second category are related to the capabilities that a FAIR data management environment should exhibit to enable and support the realization of a FAIR dataset. These tasks are categorized as ”FAIRification capabilities and resources” and include considerations such as data access, data hosting, ontology services and data sharing amongst others. These capabilities are also expected to vary depending on the level of maturity achieved or targeted.

The project examination phase also represents the target phase to employ the FAIR assessment methodology of choice to quantify the level of FAIRness exhibited by the data based on its current characteristics and environment. The assessment outcomes then help shape the necessary steps and requirements needed to achieve the desired FAIRification endpoint.

Phase 3: Assess, design, implement - then iterate

The practical part of the process centres around the FAIRification cycle, which consists of three separate stages: assessment, design and implementation. This phase typically

consists of multiple FAIRification cycles applied in an iterative fashion. Each FAIRification cycle focuses on a single FAIRification goal. We observed that three-month cycles provided the balance that delivered the best results. Three months allows for sufficient time for small, cross-functional teams to deliver observable improvements towards the overall goal, whilst balancing the need for regular validation through assessment. Three-month cycles also ensured a tight focus in work planning, mitigating the risk of insufficiently well-defined implementation tasks that we observed with longer cycles.
An assessment step sits both at the start and the end of each cycle, with the output assessment of one cycle usually serving as input to the next one. For the first cycle, the assessment will usually have been completed as part of the project examination phase. During the design stage, concrete steps from the FAIRification template are identified to achieve the FAIRification goal identified for this cycle. These steps form the FAIRification workplan to be realised during the implementation stage.

Phase 4: Review against the goals

In this final phase, the cumulative outputs of all the FAIRification processes are reviewed against the initial project goals to assess the overall success of the process. We shaped this stage in a fashion similar to the peer review process employed by academic publications, with individuals not directly involved in the practical implementation work but familiar with the overall data reviewing of the outcomes of the FAIRification work against the initial goals. We identified the need for this because it sets a clear endpoint for the FAIRification work as well as providing independent feedback on the effectiveness of the tasks. Without the review phase, there is a danger of work continuing beyond the point where the benefits to the project justify the continued expenditure of resources.

The FAIRification Template

The FAIRification Template, shown in Figure 4, implements the FAIRification Process by providing a set of clear and distinct steps for the implementation stage in the FAIRification cycle phase of the process. The template consists of eight steps relating to data hosting, such as data retrieval and data versioning, to data representation and format, such as applying data standards and vocabulary alignment, and to data content, such as identifier minting and annotation with vocabularies. A more detailed explanation of each step in the template can be found in Supplementary Table 2.

The FAIRification Workplan

The FAIRification Workplan is a specific design and implementation plan customised to an individual project by selecting the template elements required to achieve the intended FAIRification goals according to the project examination. An example of a FAIRification workplan is shown in Supplementary Figure 1.
For many of the steps in the workplan, solutions may already exist, in the shape of FAIR Cookbook[8,9] recipes (https://faircookbook.elixir-europe.org), which serve as a guide. Supplementary Table 2 provides links to the recipes that address specific implementation considerations. Once the workplan has been designed, it is put into action within the agreed cycle time frame by either following the steps from an existing recipe or implementing a

novel solution, which should then ideally be documented as a new Cookbook recipe, helping build content for others in future who face similar issues.

Discussion

Whilst developing our FAIRification process, we learned three key lessons and summarised these into take-home messages, in Box 1.

1. **Focus on incrementally achievable targets.** Projects approach the FAIR principles in different orders and risk overdoing. Instead, we focus on achieving the elements of FAIRness that matter most to the needs of the project to reach a balanced “FAIR enough” status.
2. **Tailor the FAIRification process to individual needs.** Projects have different needs, even when the underlying data, capability and resource requirements appear to be quite similar. Customising the relevant template elements allows the formation of a coherent workplan.
3. **Assemble a multi-disciplinary team.** Projects are often multi-partner, international and distributed in nature, with datasets of different provenance and source. A successful FAIRification process starts with bringing together diverse teams that include the data owners as well as professionals who can tackle the legal, curational and technical infrastructure aspects.

Box 1: Take-home messages

Developing the FAIRification framework was an iterative process: a journey we tailored to the IMI projects’ real data needs and scenarios[4]. To clarify the fundamental steps of the FAIRification process, we built on the prior state of the art described by Jacobsen and colleagues[10], work undertaken by the EHDEN project (http://www.imi.europa.eu/projects-results/project-factsheets/ehden), an IMI sibling project in the area of health data research, and by the Pistoia Alliance with the FAIRtoolkit (https://fairtoolkit.pistoiaalliance.org/), refining and expanding as needed to shape it to fit specific requirements, while remaining compatible with community practice, for instance as outlined by the GO-FAIR initiative[11]. Guidance on suitable criteria for evaluating the costs and benefits of performing FAIRification tasks on any dataset or project, in particular for the retrospective FAIRification of existing data, is discussed elsewhere[6] and lies outside the scope of the FAIRification framework.

The FAIRification process was initially presented as a linear workflow focusing on the tasks that are involved in the FAIRification of a dataset. It progressively evolved into the current process with a core iterative component to reflect the cyclic nature of improving a dataset’s FAIRness and maturity levels as well as evolving FAIRification capabilities. Another example is the composition of the FAIRification framework, which initially consisted of a single level with the elements that are now part of the template. The abstraction of the process took a step back from the implementation considerations inherent in the template, while the development of a workplan from the template emerged as a natural consequence of the design and implementation phases of the FAIRification cycle, which involved picking the specific steps and sub-steps required to achieve the FAIRification goal. The distinction between template and workplan also helped to communicate to data owners that there is a

clear path from FAIRification goals to the tasks and steps required to reach a higher level of maturity.

The successful development and implementation of a given FAIRification workplan are only possible through the assembly of a multidisciplinary team. Required roles and skills depend of course on the nature of the project and FAIRification goals but can include data managers or stewards, ontologists, curators, data scientists, software developers, system administrators and project managers. In particular, the implementation of FAIR solutions often requires technical skills such as ontology engineering, building *"*extract, transform and load" (ETL) procedures or designing FAIR-compliant data hosting solutions. Identifying the most suitable areas for improvements and thus the definition of FAIRification goals requires an in-depth understanding of the structure and content of the data, its representation and hosting requirements. This can only be achieved through close interaction with the data and a complete understanding of the lifecycle of the dataset.

A number of our FAIRification efforts were hampered and delayed by difficulties to set up appropriate legal agreements to arrange data access due to restrictive and complex data sharing policies and by insufficient buy-in from data owners due to lack of personnel, knowledge and budget available for legacy projects. Data licensing and data availability are key elements of the FAIR principles and should therefore always receive due consideration. This experience echoes that of IMI eTRIKS (http://www.imi.europa.eu/projects-results/project-factsheets/etriks), which reported similar issues[12].

The early stages of developing the FAIRification process made use of an extensively documented previous study[13] and four very different pilot projects to test the initial steps and assumptions of the process. One pilot (http://w3id.org/faircookbook/FCB045) project dataset used, the ReSOLUTE (Research Empowerment on Solute Carriers, http://www.imi.europa.eu/projects-results/project-factsheets/resolute) transcriptomics dataset, was publicly available in the SRA archive with additional information about the cell cultures and cell lines provided in separate PDF files, which is hard to efficiently extract and reuse. To improve the data findability, curators developed ETL procedures that mined experimental details from PDF files and enriched metadata about cell cultures. These sample descriptions, annotated with ontology terms enabling ontology-powered searching, were validated against the MINSEQE minimum metadata checklist[14] (https://fairsharing.org/FAIRsharing.a55z32) in order to ensure that they met broader community standards, and submitted to the Biosamples database[15]. The cell line sample metadata was also cross-referenced to the corresponding Cellosaurus[16] ID to link them to the Cellosaurus knowledge base for easier data interpretation.

This yielded a number of learning points. Firstly, retrospectively achieving compliance with "community reporting guidelines" (see for example https://fairsharing.org/search?recordType=reporting_guideline) can be challenging owing to the need to interpret a narrative rather than being able to access machine-actionable data[17]. Second, some leading archives rely on earlier-generation models which provide little support for ontologies and semantic annotations, which hampers interoperability and findability. Finally, engagement from the data owners is essential to maximise FAIRification gains.

In addition to the direct project interactions, some of the pharma partners in FAIRplus also trialled the framework through a range of specific FAIRification objectives, which provides some evidence for the broad applicability of our FAIRification process. One use case revolved around enhancing interoperability by developing an application ontology to integrate multi-omics data from independent sources, a challenge faced by the

Boehringer-Ingelheim partner. Their proposed solution made extensive use of a specific FAIR Cookbook recipe (https://w3id.org/faircookbook/FCB023). Another challenge, from the AstraZeneca partner, focused on a FAIRification goal looking at expressing “allowed data use” in a way compatible with a DCAT-based data catalogue to increase findability and reusability. The output of the work yielded a new content type in the FAIR Cookbook (https://w3id.org/faircookbook/FCB035).

Although much progress has been made to make the FAIR principles tangible by providing concrete examples, there is neither a single one-size-fits-all approach to realising FAIR in the life sciences in general nor a community-wide consensus on a FAIR representation of any given data type. The FAIRification framework provides a valuable tool to guide FAIRification efforts in a range of communities and for a variety of data types. There already exists a wide range of tools, standards, indicators and measures developed to improve data FAIRification practice, such as FAIRness assessment frameworks proposed by the RDA[18] or FAIRsFAIR[19], the Data Use Ontology (DUO)[20] standard for encoding data reuse conditions or the biosciences specific resource markup framework, Bioschemas[21]. The framework is agnostic of any specific indicators or implementation and any of these can be plugged into the framework in the relevant places.

The successful application of the framework in both exemplar projects and its integration into the working processes of several pharma partners demonstrates its broad applicability to any life sciences data. Supported by an active and knowledgeable community passionate about the importance of bringing FAIR to the forefront of scientific data management, it should serve as a guide to anyone looking to address FAIRification challenges.

Methods

Incremental framework development

The development of the FAIRification framework was an iterative process. It was developed over the course of two years, starting with a set of four pilot projects whose FAIRification served to establish the basic structure and elements of the process. This was then refined over several iterations, using a wide range of IMI projects as well as FAIRification use cases elicited from EFPIA partners, to establish the framework described above.

Both the pilot projects and subsequent projects were selected from the full pool of IMI projects through a formal process. The details of this process and how it was established are discussed elsewhere[6]. Once selected, projects were passed on to cross-disciplinary working groups who worked with the data owners to set FAIRification goals and progress the project through the steps and stages of the framework.

The FAIRification template was developed to accommodate a wide range of projects and data types. The steps and sub-steps of the FAIRification template were refined from data FAIRification efforts and experience in a wide range of contexts and domains and from the prior experience of cross-disciplinary task teams within the FAIRplus project (see below). Elements of the template are more relevant to some areas than others but overall, the template can be applied and customised to any type of project, rather than being applicable to only very specific data types, such as healthcare or clinical trial data.

Cross-disciplinary task teams

The practical work executed during the FAIRplus project was carried out by cross-disciplinary teams, referred to as 'Squads' (to borrow the terminology of the Spotify model described in https://www.atlassian.com/agile/agile-at-scale/spotify), assembled to provide the expertise required for a given task. The working practices and methodology of these squad teams were iteratively refined over 2 years, and a report on this process is in preparation. The personnel were recruited to squads from across all work packages, allowing the incorporation of specialist knowledge, and fostering information exchange within the project. Besides this base composition, other floating team members were recruited to address specific and arising needs in the short term, allowing a flexible and tailored response. Squad representatives engaged early in project discussions between IMI project representatives and FAIRplus triage staff, allowing a preview of the types of data and issues that may be coming through the onboarding pipeline, and determining whether potential areas of work could improve the content of the FAIR Cookbook. This also allowed a level of expectation management regarding the distribution of work for the actual FAIRification tasks between project representatives and FAIRplus personnel, as well as the development of a collaborative relationship with external project representatives. During these exchanges, the squads would engage with project representatives to identify tasks and goals that were realistically achievable in the given time frames, routinely being of roughly 3 months total duration.

Validation process

In parallel to the development of the FAIRifiation framework, we also developed a FAIR DataSet Maturity (FAIR-DSM) Model (https://fairplus.github.io/Data-Maturity/). This allowed us to assess the maturity of the datasets used to validate the FAIRification process prior to and following any FAIRification work. In the initial stages of the framework development, we made use of existing approaches including the RDA indicators[18] and the FAIRsFAIR indicators[19] to evaluate FAIRifiation improvements. While these alternative models demonstrated satisfactory results, they generally treat each element or principle of FAIR as a stand-alone element. The FAIR-DSM on the other hand evaluates a dataset as a whole, providing a more balanced view of its overall maturity in terms of content, representation and hosting.

The FAIR-DSM is described in detail elsewhere[5] but briefly, it consists of 5 maturity levels characterised by increasing requirements across all aspects of FAIR, plus a reference level, referred to as "level 0", representing a state of data that is missing most or all fundamental FAIR requirements. The model considers 3 categories of requirements: content-related requirements; representation and format requirements; and hosting environment capabilities. In order to conform to a given level of the model, a dataset needs to fulfil a set of indicators covering the requirements for each of the 3 categories at this level. Figure 5 provides a summary description and perspective for each level.

Ancillary materials

The FAIR Cookbook

The learnings and insights gained from the efforts of the FAIRplus project were distilled into individual “recipes” making up the FAIR Cookbook. This specific practical guidance is intended to provide concrete solutions to common FAIR data problems. The FAIR Cookbook is available at https://faircookbook.elixir-europe.org/.

The FAIR Wizard, mining the FAIR Cookbook

We recognize that providing guidance that leads to a specific implementation of the FAIRification process is still an expert activity. To provide support for those who are newer to FAIR implementation, we have begun developing a “FAIR Wizard” (https://www.ebi.ac.uk/ait/fair-wizard/) to facilitate the work of project managers and data scientists in identifying FAIRification goals, examining projects and developing FAIRification solutions. The wizard as a whole is effectively based on the FAIRification template, with the output provided to a user representing a skeleton FAIRification workplan.
The FAIR wizard collects FAIRification goals from datasets that we worked with and the knowledge consolidated in the FAIR Cookbook in the form of curated solutions for the common use cases, which can be reused directly. It also assists people in identifying their own use cases through a series of questions and FAIR assessments and proposes solutions accordingly.
The FAIR wizard utilises FAIRification resources developed by this project and other platforms, suggests FAIRification materials based on the FAIRification requirements, and designs FAIRification solutions for data owners, data stewards and other people involved in FAIRification.

The IMI Data Catalog

All datasets engaged during the establishment and validation of the FAIRification framework were included in the IMI Data Catalog[22] (https://datacatalog.elixir-luxembourg.org/) hosted by ELIXIR Luxembourg. Dataset entries include information on the maturity level of the dataset before and after FAIRification efforts as well as key metadata about the project, experimental process and dataset.

Data Availability

- IMI Data Catalog: https://datacatalog.elixir-luxembourg.org/
- EBI repositories
    - ReSOLUTE data in BioSamples: https://www.ebi.ac.uk/biosamples/samples?filter=attr:project:RESOLUTE
    - CARE
        - Pre-FAIRification: https://zenodo.org/record/5872683#.YvZff-xBwbk
        - FAIRified data in ChEMBL: http://dx.doi.org/10.6019/CHEMBL4651402
- FAIRification results: https://fairplus.github.io/fairification-results/

Code Availability

- FAIRplus Github organisation: https://github.com/FAIRplus
- FAIR Maturity Model: https://github.com/FAIRplus/Data-Maturity (MIT license)
- FAIR Wizard: https://github.com/FAIRplus/FAIR_wizard (Apache-2.0)
- FAIR Cookbook: https://github.com/FAIRplus/the-fair-cookbook (CC BY 4.0)
- IMI Data Catalog: https://github.com/FAIRplus/imi-data-catalogue (AGPL-3.0 for code, CC BY-NC-SA 4.0 for data)

Acknowledgement

The authors would like to thank all members of the FAIRplus consortium for their contributions to discussions and the development of the FAIRification framework, in particular the members of the FAIRplus "squad" teams. This work and the authors were funded by FAIRplus (IMI 802750).

Author Contributions

All authors contributed to the conception and refinement of the framework. DW, SAS and TB wrote the manuscript. All authors critically revised the paper for intellectual content and approved the final version of the manuscript.

Competing Interests

SAS is Honorary Academic Editor of *Scientific Data* and PRS is a member of the *Scientific Data* Senior Editorial Board.

References

1. The Lancet Digital Health. Transparency during global health emergencies. *Lancet Digit. Health* **2**, e441 (2020).
2. Badker, R. *et al.* Challenges in reported COVID-19 data: best practices and recommendations for future epidemics. *BMJ Glob. Health* **6**, e005542 (2021).
3. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
4. Burdett, T. *et al.* FAIRplus: D3.3 Report on IMI projects for data types and current technical solutions. *Zenodo* https://doi.org/10.5281/zenodo.4428721 (2021).
5. Emam, Ibrahim *et al.* FAIRplus D2.6 FAIR Data Set Maturity model. *Zenodo*

https://doi.org/10.5281/zenodo.7464523 (2022).

6. Alharbi, E. *et al.* Selection of data sets for FAIRification in drug discovery and development: Which, why, and how? *Drug Discov. Today* **27**, 2080–2085 (2022).
7. Gaulton, A. *et al.* The ChEMBL database in 2017. *Nucleic Acids Res.* **45**, D945–D954 (2017).
8. Rocca-Serra, P. *et al.* The FAIR Cookbook - the essential resource for and by FAIR doers. Preprint at https://doi.org/10.5281/zenodo.7156792 (2022).
9. Rocca-Serra, P. *et al.* D2.1 FAIR Cookbook. *Zenodo* https://doi.org/10.5281/zenodo.6783564 (2022).
10. Jacobsen, A. *et al.* A Generic Workflow for the Data FAIRification Process. *Data Intell.* **2**, 56–65 (2020).
11. Sustkova, H. P. *et al.* FAIR Convergence Matrix: Optimizing the Reuse of Existing FAIR-Related Resources. *Data Intell.* **2**, 158–170 (2020).
12. Gu, W., Hasan, S., Rocca-Serra, P. & Satagopam, V. P. Road to effective data curation for translational research. *Drug Discov. Today* **26**, 626–630 (2021).
13. Rocca-Serra, P. & Sansone, S.-A. Experiment design driven FAIRification of omics data matrices, an exemplar. *Sci. Data* **6**, 271 (2019).
14. Brazma, Alvis *et al.* MINSEQE: Minimum Information about a high-throughput Nucleotide SeQuencing Experiment - a proposal for standards in functional genomic data reporting. *Zenodo* https://doi.org/10.5281/zenodo.5706412 (2012).
15. Courtot, M., Gupta, D., Liyanage, I., Xu, F. & Burdett, T. BioSamples database: FAIRer samples metadata to accelerate research data management. *Nucleic Acids Res.* **50**, D1500–D1507 (2022).
16. Bairoch, A. The Cellosaurus, a Cell-Line Knowledge Resource. *J. Biomol. Tech.* **29**, 25–38 (2018).
17. Batista, D., Gonzalez-Beltran, A., Sansone, S.-A. & Rocca-Serra, P. Machine actionable metadata models. *Sci. Data* **9**, 592 (2022).
18. Research Data Alliance FAIR Data Maturity Model Working Group. FAIR Data

Maturity Model: specification and guidelines. *Zenodo* https://doi.org/10.15497/rda00050 (2020).

19. Devaraju, A. *et al.* FAIRsFAIR Data Object Assessment Metrics. *Zenodo* https://doi.org/10.5281/zenodo.4081213 (2020).

20. Lawson, J. *et al.* The Data Use Ontology to streamline responsible access to human biomedical datasets. *Cell Genomics* **1**, 100028 (2021).

21. Gray, A. J. G., Goble, C. & Jimenez, R. Bioschemas: From Potato Salad to Protein Annotation. *Int. Semantic Web Conf. Posters Demos Ind. Tracks* (2017).

22. Welter, D. *et al.* The Translational Data Catalog - discoverable biomedical datasets. Preprint at https://doi.org/10.5281/zenodo.7157285 (2022).

Figure 1: The three components of the FAIRification Framework: the conceptual FAIRification Process, the FAIRification Template covering all aspects of FAIRification and the FAIRification Workplan as a single tailored implementation guide.

Figure 2: FAIRification Process composed of four distinct phases. This is a reduced version of the process diagram. The full version, with additional explanatory text, is available in Supplementary Figure 2.

Figure 3: Dataset maturity levels for 17 projects before and after passing through the FAIRification Process. Maturity levels are broken down into representation-related, content-related and hosting-related maturity. The assessments were performed using the Dataset Maturity (DSM) model indicators developed by FAIRplus. It is important to note that maturity improvements should not be compared between projects as they are highly dependent on the specific characteristics of each dataset and the chosen FAIRification goals.

Figure 4: The FAIRification template steps

Figure 5: FAIR-DSM levels.

1. FAIRification Process

Define FAIRification Goal → Project Examination → Iterative FAIRification Cycles → Post-FAIRification Review

2. FAIRification Template

Get the Data → Model The Domain → Select an Identifier Scheme → Apply a Data Standard → Choose Data Vocabulary → Transform Data for Interoperability → Host Data → Share Data

3. FAIRification Workplan

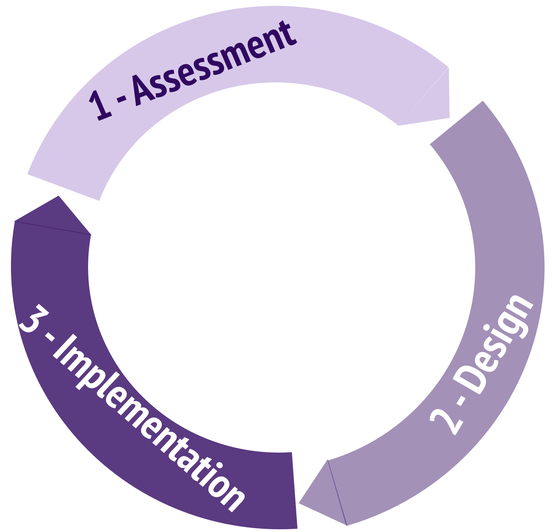
Design Decisions	Identifier Strategies	Metadata Strategies	Ontology Strategies	Data Sharing Strategies
Implementation Phase	Task 1: DONE	Task 3: IN PROGRESS	Task 4: IN PROGRESS	Task 6: DONE
	Task 2: TO DO		Task 5: TO DO	

Define FAIRification Goal

Project Examination

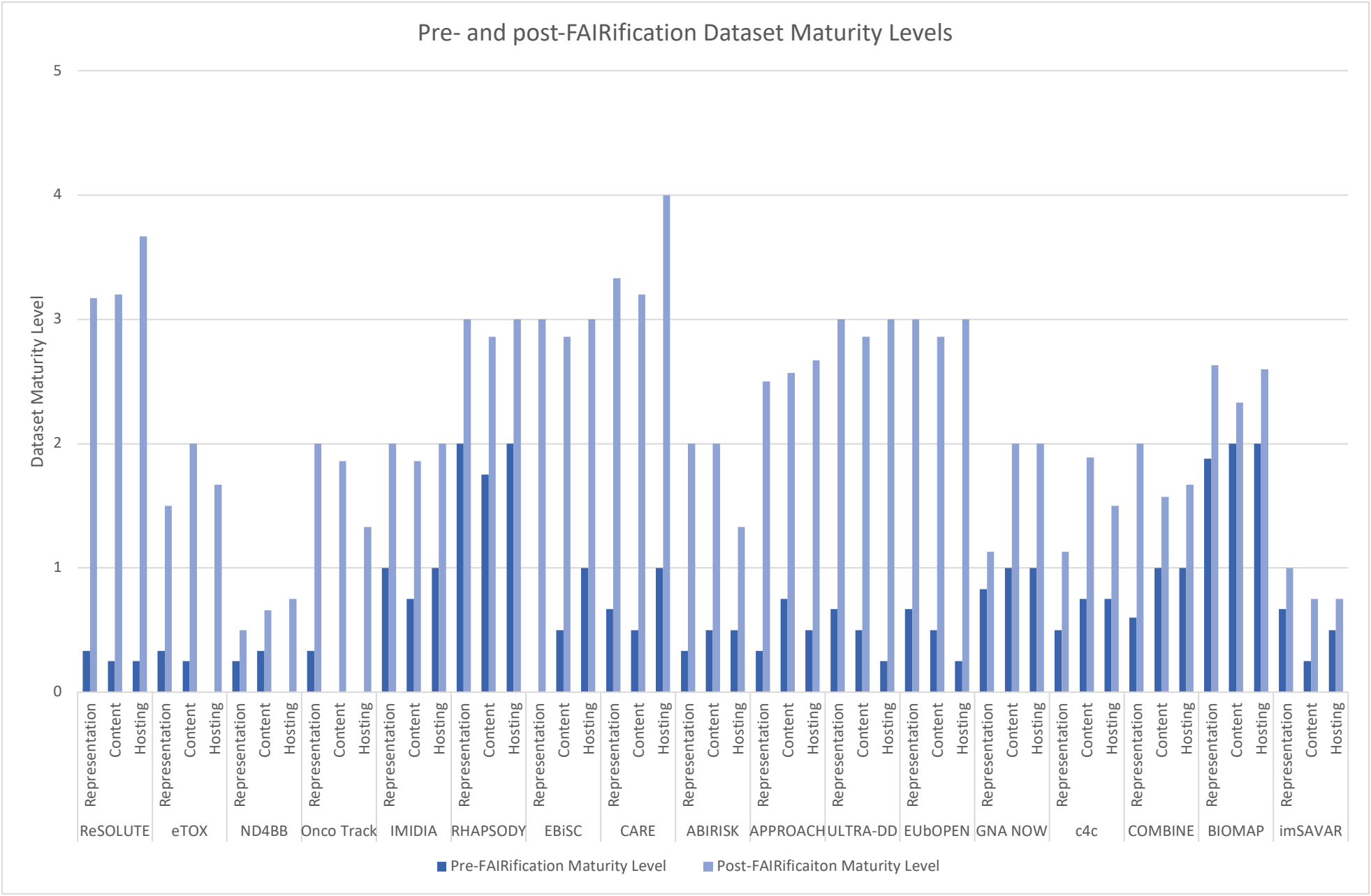
- Identify Data Requirements
- Identify FAIRification Capabilities & Resources
- Produce the FAIRification Backlog

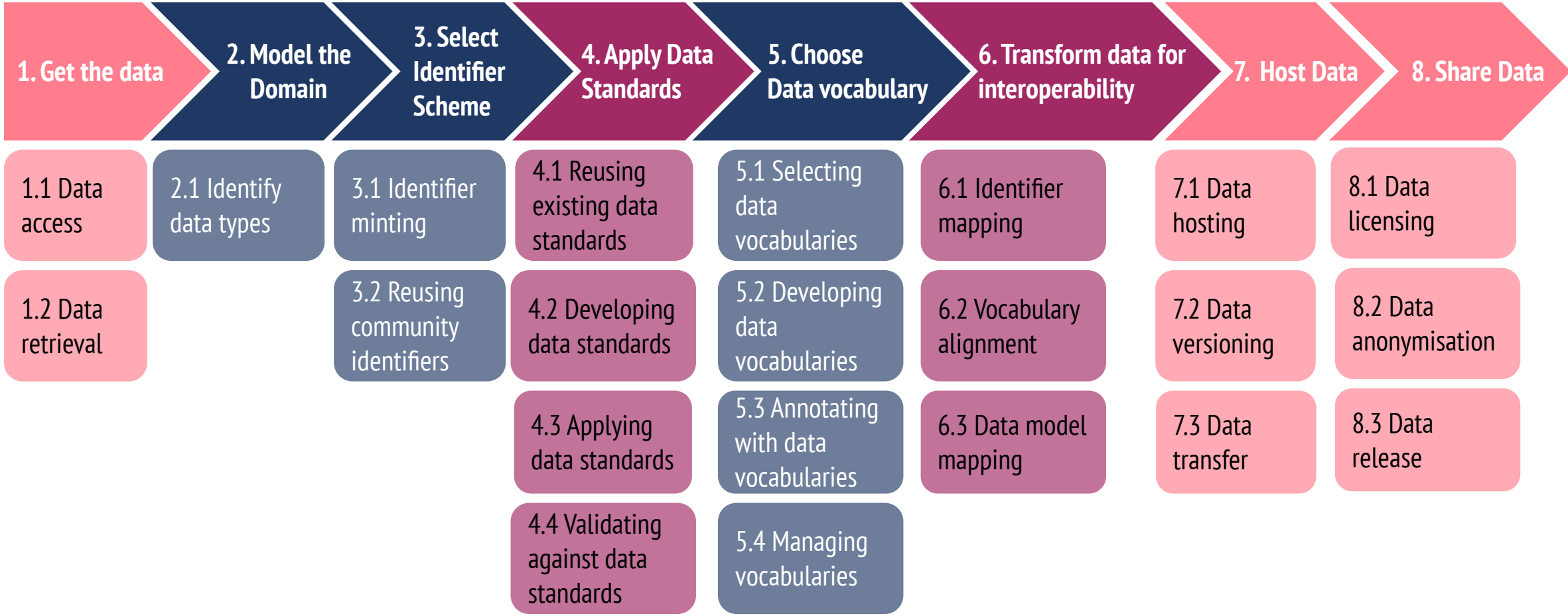
Iterative FAIRification Cycles

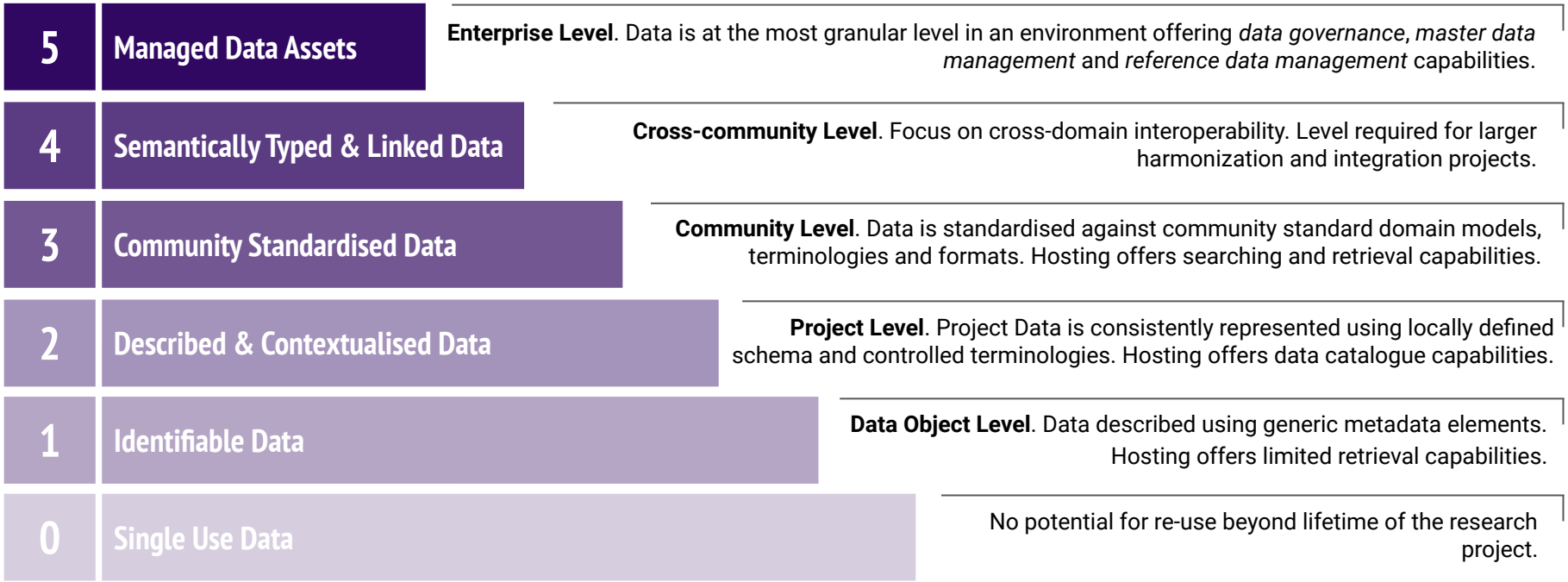


Post-FAIRification Review

- Disseminate learnings


Pre- and post-FAIRification Dataset Maturity Levels
Dataset Maturity Level
0
1
2
3
4
5
Representation
Content
Hosting
ReSOLUTE
eTOX
ND4BB
Onco Track
IMIDIA
RHAPSODY
EBiSC
CARE
ABIRISK
APPROACH
ULTRA-DD
EUbOPEN
GNA NOW
c4c
COMBINE
BIOMAP
imSAVAR
Pre-FAIRification Maturity Level
Post-FAIRificaiton Maturity Level


1. Get the data
2. Model the Domain
3. Select Identifier Scheme
4. Apply Data Standards
5. Choose Data vocabulary
6. Transform data for interoperability
7. Host Data
8. Share Data
1.1 Data access
1.2 Data retrieval
2.1 Identify data types
3.1 Identifier minting
3.2 Reusing community identifiers
4.1 Reusing existing data standards
4.2 Developing data standards
4.3 Applying data standards
4.4 Validating against data standards
5.1 Selecting data vocabularies
5.2 Developing data vocabularies
5.3 Annotating with data vocabularies
5.4 Managing vocabularies
6.1 Identifier mapping
6.2 Vocabulary alignment
6.3 Data model mapping
7.1 Data hosting
7.2 Data versioning
7.3 Data transfer
8.1 Data licensing
8.2 Data anonymisation
8.3 Data release
FAIR Dataset Maturity Model Category of Requirements
Content related
Representation & Format
Hosting Environment Capabilities


5
Managed Data Assets
Enterprise Level. Data is at the most granular level in an environment offering data governance, master data management and reference data management capabilities.
4
Semantically Typed & Linked Data
Cross-community Level. Focus on cross-domain interoperability. Level required for larger harmonization and integration projects.
3
Community Standardised Data
Community Level. Data is standardised against community standard domain models, terminologies and formats. Hosting offers searching and retrieval capabilities.
2
Described & Contextualised Data
Project Level. Project Data is consistently represented using locally defined schema and controlled terminologies. Hosting offers data catalogue capabilities.
1
Identifiable Data
Data Object Level. Data described using generic metadata elements. Hosting offers limited retrieval capabilities.
0
Single Use Data
No potential for re-use beyond lifetime of the research project.

Supplementary Table 1 - IMI projects that interacted with FAIRplus

Project	Engagement phase	Data types	FAIRification goal(s)	Outputs/recipes
Onco Track	Pilot project	Patient-derived samples (oncology)	Convert the Onco Track sample metadata to a structured and consistent data format, improves the findability, interoperability, and reusability of the metadata	FCB044
ND4BB	Pilot project	in-vitro data on compound properties for known antibiotics	Creating a FAIR & machine-readable data set of the AMR database	FCB043
eTOX	Pilot project	Chemical compounds, toxicology assays	Semantic markup to reduce free text descriptors	FCB042
ReSOLUTE	Pilot project	Transcriptomics, proteomics, metabolomics	1. Deposition to a public repository and compliance to community standard (MINSEQE) 2. Conversion from proprietary to open format	FCB045; FCB029
IMIDIA	FAIRification process v2.0	Clinical data, transcriptomics	1. Identify gaps in current metadata annotations and pick the best ontologies to fill them 2. Make metadata findable/searchable	n/a
RHAPSODY	FAIRification process v2.0	Clinical data, transcriptomics	1. Identify gaps in current metadata annotations and pick the best ontologies to fill them 2. Make metadata findable/searchable	n/a
EBiSC I & II	FAIRification process v2.1	Cell line metadata, genomics	EBiSC seeks to make specialised cell lines as findable as possible for its users, based on a selected (small) set of relevant descriptors.	n/a
APPROACH	FAIRification	Clinical trial data,	1. Map the metadata parameters (data dictionary) to	FCB078;

	process v2.1	imaging, biomarkers	appropriate domain-relevant ontologies and standards to enable applying to data catalogues and repositories to make the data more findable. 2. Provide advice and information to the consortium members so they can decide on the type of licensing for publicly sharing the data and clarifying the possible reuse of the data.	FCB025
ABIRISK	FAIRification process v2.1	Clinical trial data	1. Map the data dictionary to CDISC and appropriate domain-relevant ontologies to facilitate data interoperability and enable sharing of metadata in data catalogues and repositories to make the data more findable. 2. Provide advice and information to the consortium members so they can decide on the type of licensing for publicly sharing the data and clarifying the possible reuse of the data.	FCB078; FCB025
CARE	FAIRification process v2.1	Compound and bioassay data	To publish data in open archives and comply with community data standards so that other researchers can find and reuse the compound and bioassay data.	FCB057
ULTRA-DD	FAIRification process v2.1	High-content screen data	Promote public access, data dissemination and sharing of project datasets	n/a
EUbOPEN	FAIRification process v2.1	High-content screen data, bio-imaging data	EUbOPEN seeks to make multimodal chemical biology assays as findable as possible to facilitate dataset discovery based on a small number of search criteria	FCB067
COMBINE	FAIRification process current version	Bioassay protocol data	Composition of an application ontology to aid in reproducibility of in-vivo bioassay experiments	FCB023; https://github.com/Fraunhofer-ITMP/bpo
c4c	FAIRification process	Clinical trial metadata, eCRFs	Study- and protocol-level additional (meta)data (such as in/exclusion criteria) required alongside the CRF data	In progress

	current version - work ongoing		dictionary to make the overall trial data more findable in relation to this information: ● Define & refine list of variables to be collected ● Represent protocol-level additional (meta)data in a complementary data model ● Define extraction processes for populating variables of interest	
BIOMAP	FAIRification process current version - work ongoing	Clinical trial data, omics data	1. FINDABILITY Improve findability of project metadata for external researcher through publication of metadata in the IMI Data Catalog 2. INTEROPERABILITY (Re)Align the data glossary with the OMOP community standard in order to improve the data's interoperability with other OMOP datasets 3. REUSABILITY Define and implement QC policies/best practice to ensure that data files can be reached from patient metadata and that data files are in the correct format as defined by the metadata	In progress
GNA NOW	FAIRification process current version - work ongoing	in-vivo and in-vitro efficacy data	Standardization and development of workflows involving data archiving process for terminated sub-projects	In progress
imSAVAR	FAIRification process current version - work ongoing	Omics data	1. Create a data dictionary that is harmonised across species as well as being consistent with the prospective metadata collection form of imSAVAR in order to facilitate data reuse across the imSAVAR project. Includes harmonising/mapping terms against a CV where possible 2. Design a metadata template to capture study/protocol-level contextual metadata about the dataset, including antibodies, preparation methods and analysis workflows, in order to improve reusability and data integration	In progress

			3. Suggest data usage conditions to data owner and provide machine-readable sample encoding of conditions for Data Catalog metadata in order to illustrate machine actionable metadata	
ESCulab	FAIRification process current version - work ongoing	Compound and bioassay data	1. Improve the searchability of the data for current and future users for analysis and reuse by enhancing and structuring existing metadata. 2. Accessibility: Providing metadata and exposing the data 3. Interoperability and Reusability: Improving future interoperability for the project after the timeline of the project 4. Reusability: Enable privacy preserving analysis of the data with third parties	In progress
U-BIOPRED	FAIRification process current version - work ongoing	Omics data	*To be determined*	*To be determined*
eTRANSAFE	FAIRification process current version - work ongoing	Chemical toxicity prediction tool	*To be determined*	*To be determined*

Supplementary Table 2 - steps of the FAIRification template

Capabilities domain	**General FAIRification step**	**FAIRification sub-step**	**Description**	**Related FAIR Cookbook recipes**
Hosting environment capabilities	1. Get the data	1.1 Data access	Considerations relating to how data is accessed, eg through APIs, via controlled access	FCB014, FCB015, FCB073
		1.2 Data retrieval	Considerations relating to data retrieval, eg query language, results representation and exporting capabilities	FCB040, FCB046, FCB060, FCB070
Content-related capabilities	2. Model the domain	2.1 Identify data types	Data type identification informs the selection of appropriate data standards, ontologies and target repositories	FCB027, FCB057
	3. Select the identifier scheme	3.1 Identifier minting	How to create unique, persistent and resolvable identifiers	FCB006, FCB007, FCB008, FCB077
		3.2 Reusing community identifiers	How to reuse existing identifiers in a dataset	FCB016, FCB017
Representation & format capabilities	4. Apply data standards	4.1 Reusing existing data standards	How to reuse existing data standards	FCB025
		4.2 Developing data standards	How to develop a new data standard if no appropriate standards exist	FCB025, FCB026, FCB027
		4.3 Applying data standards	How to apply data standards to datasets, especially retroactively	FCB025, FCB029, FCB078

		4.4 Validating against data standards	How to use validation to ensure that a dataset is compliant with a data standard	FCB028, FCB030
Content-related capabilities	5. Choose data vocabularies	5.1 Selecting data vocabularies	How to select the most appropriate vocabularies to annotate a dataset	FCB019, FCB020
		5.2 Developing data vocabularies	How to develop new vocabularies from scratch	FCB021
		5.3 Annotating with data vocabularies	How to annotate data and metadata with terms from vocabularies	FCB022, FCB023
		5.4 Managing vocabularies	How to manage vocabularies and ontologies	FCB003, FCB004, FCB005, FCB022
Representation & format capabilities	6. Transform data for interoperability	6.1 Identifier mapping	How to map between different types of equivalent identifiers	FCB016, FCB017, FCB018
		6.2 Vocabulary alignment	How to map between different equivalent vocabulary terms	FCB022
		6.3 Data model mapping	How to map equivalent concepts from different data models	FCB016, FCB031, FCB058, FCB059, FCB065
Hosting environment capabilities	7. Host your data	7.1 Data hosting	Considerations around data hosting infrastructure such as markup and search engine optimisation	FCB009, FCB010, FCB011, FCB012, FCB013, FCB047, FCB048
		7.2 Data versioning	Considerations around data versioning	FCB009, FCB036
		7.3 Data transfer	Considerations around data transfer such as file formats, repository types and checksumming	FCB014, FCB015, FCB052, FCB053
	8. Share your	8.1 Data licensing	Data licensing considerations such as	FCB032, FCB033,

	data		which license is most appropriate for a given scenario	FCB034, FCB035, FCB036
		8.2 Data anonymisation	Data anonymisation considerations	n/a - due to the complexity of this subject, incl legal ramifications, the FAIR Cookbook does not include guidance on data anonymisation
		8.3 Data release	Data release considerations such as when to release a dataset and where to release it	FCB009, FCB061, FCB067

Supplementary Figure 1 - FAIRification Workplan example for the CARE project

This figure shows an example of a real FAIRification workplan, as used for the CARE project. The top half of the workplan covers the FAIRification goal, the key findings of the project examination and the results of the pre-FAIRification maturity assessment. The middle section lists the “Design decisions” or high-level tasks required to achieve the stated FAIRification goal, grouped into logical categories. The bottom section shows the progress of the implementation phase, with concrete steps required to complete each of the tasks from the previous section. In this example, all steps are marked in green, indicating that they have been completed. Additional colour coding is available to mark tasks as To Do (orange), In Progress (blue) or At Risk (red). The latter category is used for tasks that might not be completed during the current FAIRification cycle due to circumstances that emerged after the task had been prioritised.

FAIRplus Tailored FAIRification Process - CARE - Iteration 2

1- Define FAIRification Goal

Define desired and expected outcomes of FAIRification.

To publish data in open archives and comply with community data standards so that we share identification and testing of compounds for therapeutic properties.

- **Findability**: Improve Findability of (meta)data by mapping to common data models and ontologies
- **Reusability**: Improve Reusability by encoding data reuse conditions in metadata

2- PROJECT EXAMINATION

Identify Data Requirements

Current	Projected
Consent/Licence form varies based on cohort	Unified license form across CARE
Inconsistent use of identifiers (mixture of internal and other)	Mapping to Standardisied identifiers (InchI and/or SMILEs)
Lab/biomarker results	Standardisation of IC50 values
- Variety of data types - Data in simple tabular format (csv)	Data can be downloaded in various standardised formats (XML, JSON)
Data dictionary not available	Metadata for project added from BAO and CHEMBL-preferred ontologies

Identify Data FAIRification Capabilities

Current	Projected
Data owners need to clarify data access and reuse criteria	Hosting platform selected will expose licenses for sharing rules in place

Identify Data FAIRification Resources

Current	Projected
Data currently only hosted internally on a temporary server or on Zenodo	Data shared in public resource (CHEMBL)
	Project and dataset metadata in IMI Data Catalogue

3- Assessment

Pre-FAIRification Assessment Results

DSM Maturity levels (FAIR DSM indicators v1.0) :
- Maturity level - Representation: 0.67
- Maturity level - Content: 0.5
- Maturity level - Hosting: 1.0

Overall maturity level: 0

NOTE: Ultimately, a single dataset (ChEMBL) was selected for use as an exemplar implementation.

4- Design Decisions

Design identifier strategies

Map current identifiers (mixture of internal and standard) to InChI and SMILE ids.

Design metadata strategies

Publish agreed metadata in the IMI Data Catalog (FAIRplus requirement)

Design ontology strategies

Apply ontologies/standardised terminologies to metadata

Design data sharing strategies

Support data sharing strategy by encoding data reuse criteria using DUO/ODRL

Determine suitable data reuse license

Identify most appropriate hosting platform (various public dbs, single internal repo with DAC, semi-public custom repo)

5- Implementation Phase

TASKS

Tasks LEGEND:

TO DO	*IN PROGRESS*
AT RISK	*DONE*

InChi and SMILE identifiers generated (recipe)

ChEMBL IDs assigned in each compound and assay. (Collection ID)

Data mapped to the ChEMBL submission model

Assay value split into numeric values and operators.

Tabular form for collection of assay from wide-short to tidy-long

CARE project metadata submitted to the IMI Data Catalog

ChEMBL model for general bioactivity data (recipe)

BioAssay Ontology(BAO), CHEMINF vocabulaires used for compound annotation

Introduced Cellosaurus ID for cell line (Vero-E6)

ChEMBL license used for data sharing, CC-SA 3.0

6- Assessment

Post-FAIRification Assessment Results

DSM Maturity levels (FAIR DSM indicators v1.0) :
- Maturity level - Representation: 3.43
- Maturity level - Content: 3.2
- Maturity level - Hosting: 4.0

Overall maturity level: 3

Supplementary Figure 2 - FAIRification Process diagram, full version

The FAIRification Process, composed of four distinct phases. This is an expanded version of the diagram shown in Figure 2 in the main text.

Define FAIRification Goal

Determine goal for FAIRification, in terms of desired *usability of data* that isn't currently possible

Project Examination

Examine the current state of the project with respect to *FAIRification goal*

Identify Data Requirements

Determine the ***indicators*** that will specify the actions needed to curate the data (wrt FAIR) to fulfil the expected FAIRification goal

Identify FAIRification Capabilities & Resources

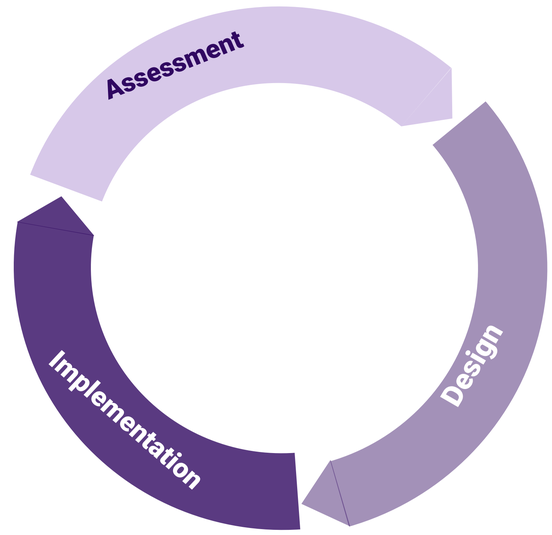
Determine data management ***capabilities*** *(e.g. data annotation, data search and indexing)* ***& resources*** (e.g. tools, databases, vocabulary services) needed to enable and support the FAIRification process

Produce the FAIRification Backlog

Define a prioritized list of target FAIRification outcomes, concentrating on the desired results for the project rather than the specifics of the solution.

Iterative FAIRification Cycles

Define and deliver *practical, achievable objectives* for a single 3 month release cycle that work towards the overall FAIRification goal



1- Assessment

Honest Evaluator to perform pre- and post-FAIRification assessments and to produce the summary report before and after every cycle

2- Design

Work with project representatives to develop a solution plan for the current cycle, which may involve developing new cookbook recipes or re-using existing ones.

3- Implementation

Execute the solution plan within the agreed time-frame. Prepare the resulting FAIRified data for the end-of cycle assessment and inform the Honest Evaluator when ready.

Post-FAIRification Review

Review *outcomes* and assess success against original goals

Disseminate learnings

Developed recipes are shared publicly in the cookbook. If applicable, update FAIRness levels in IMI catalogue. Integrate lessons learnt with other initiatives e.g. Pistoia FAIR toolkit, RDMKit