

IMI2 Project 802750 - FAIRplus  
FAIRification of IMI and EFPIA data

## WP3 – Identification of and implementation of data on sustainable data hosting platforms

### D3.6 Technical Feasibility Report

Lead contributor	Tony Burdett (1 – EMBL-EBI) tburdett@ebi.ac.uk
Other contributors	Fuqi Xu (1 – EMBL-EBI) Wei Gu (7 – UL) Nick Juty (4 – UNIMAN) Ibrahim Emam (11 – ICL) Jan-Willem Boiten (13 – LYG) Philippe Rocca-Serra (3 – University of Oxford)

Due date	30 June 2022
Delivery date	18 Jul 2022
Deliverable type	R
Dissemination level	PU

Description of Work	Version	Date
	V1.0	18 Jul 2022

## Document History

Version	Date	Description
V0.1	20 Jun 2022	First Draft
V0.2	06 Jul 2022	Final draft for comments
V0.3	14 Jul 2022	All reviewer comments addressed
V1.0	18 Jul 2022	Final Version

## Table of Contents

<b>Document History</b>	<b>2</b>
<b>Executive Summary</b>	<b>4</b>
<b>Introduction</b>	<b>4</b>
<b>Overall Approach</b>	<b>5</b>
<b>Feasibility and Requirements</b>	<b>7</b>
FAIRification processes	7
Impact	7
Creation	8
Application	8
Evaluation	9
Requirements	10
Summary of Feasibility Analysis	10
FAIR assessment	10
Impact	10
Creation	10
Application	11
Evaluation	11
Requirements	12
Summary of Feasibility Analysis	12
FAIR Cookbook	12
Impact	12
Creation	13
Application	13
Evaluation	14
Requirements	15
Summary of Feasibility Analysis	15
FAIR tooling and data hosting	16

Impact	16
Creation	16
Application	17
Evaluation	18
Requirements	19
Summary of Feasibility Analysis	19
<b>Discussion</b>	<b>19</b>
<b>Conclusion</b>	<b>20</b>
<b>Appendix A - Evolution of the FAIRplus FAIRification Process</b>	<b>21</b>

# 1. Executive Summary

FAIRplus has produced four types of technical solutions:

- 1. FAIRification processes**, and adoption of a FAIRification framework
- 2. FAIR assessment** through adoption of the FAIR-DSM maturity model
- 3. FAIR Cookbook** and application of recipes
- 4. FAIR tooling and data hosting**

This deliverable report describes how the FAIRplus approach to FAIRification, through a co-production model utilising “squad teams” that run in 3 month cycles and provides frequent checkpoints for evaluating technical feasibility, has ensured technical feasibility of project outputs in each of these four impact areas. This deliverable report described, for each of the four key impact areas, a brief description of the creation of technical solutions, their application to IMI projects and an overview of their evaluation. Based on this evaluation, we quantify the ease with which solutions can be applied and any requirements that may need to be in place prior to adoption of technical solutions from FAIRplus in future projects.

## 2. Introduction

The goal of work package 3 is the technical delivery of FAIRification, building on the work of work package 1 and 2, and leveraging, enhancing and extending existing FAIRification tools to deliver annotated datasets. We have previously defined our approach to the development of technical solutions (see D3.3, Report on IMI projects for data types and current technical solutions). As described in D3.3, our approach is data- and use-case driven and organised in five stages as follows:

**Stage 1: Sourcing of data types and corresponding IMI projects**

**Stage 2: Data type prioritisation based on industrial and academic impact**

**Stage 3: Development of technical solutions for FAIRification, tailored for specific IMI project FAIRification**

**Stage 4: FAIR assessment for evaluation of the technical solution**

**Stage 5: Documentation, recipe generation and generalisation of technical solutions to data types**

An early insight of WP3 was to identify that technical delivery requires early and frequent validation of solutions. WP3 has therefore sought to establish the technical feasibility of FAIRification solutions produced by FAIRplus through the “squads” approach to FAIRification (see D2.2, “BYOD guidelines<sup>1</sup>”). Working alongside squads, WP3 has followed an incremental and iterative approach to technical delivery, testing early solutions and prototypes across as wide a range of IMI and EFPIA projects as possible, and using regular (roughly once every three months) checkpoints (BYODs) for evaluation and validation of technical solutions. This iterative, squad-led approach was designed early on to ensure technical feasibility could be guaranteed for any

---

<sup>1</sup> <https://zenodo.org/record/6806341#.Ys2qsqjMI2x>

proposed solution.

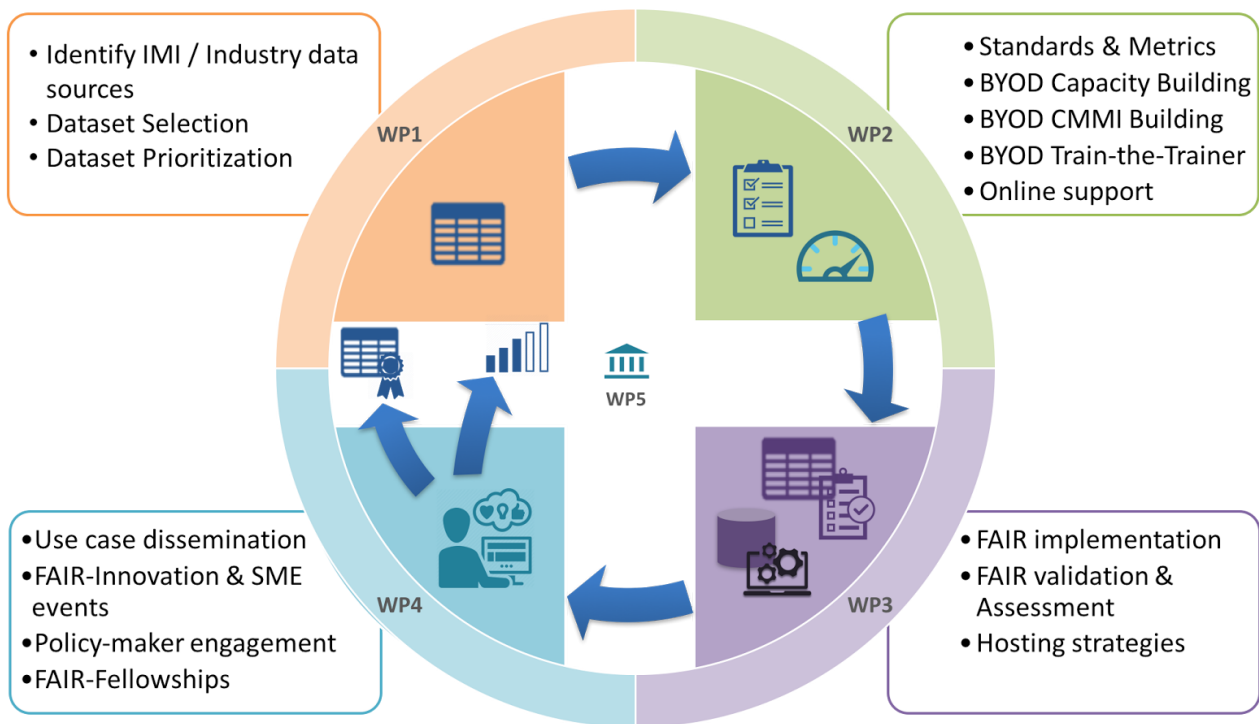
FAIRification solutions from FAIRplus are broadly categorised into the following four impact areas:

- 1. FAIRification processes**, and adoption of a FAIRification framework
- 2. FAIR assessment** through adoption of the FAIR-DSM maturity model
- 3. FAIR Cookbook** and application of recipes
- 4. FAIR tooling and data hosting**

In this report, for each of these four impact areas, we highlight our approach to validation of technical feasibility, define exemplar projects that have successfully adopted recommended technical solutions, and indicate any requirements that may need to be in place prior to adoption of technical solutions from FAIRplus in future projects.

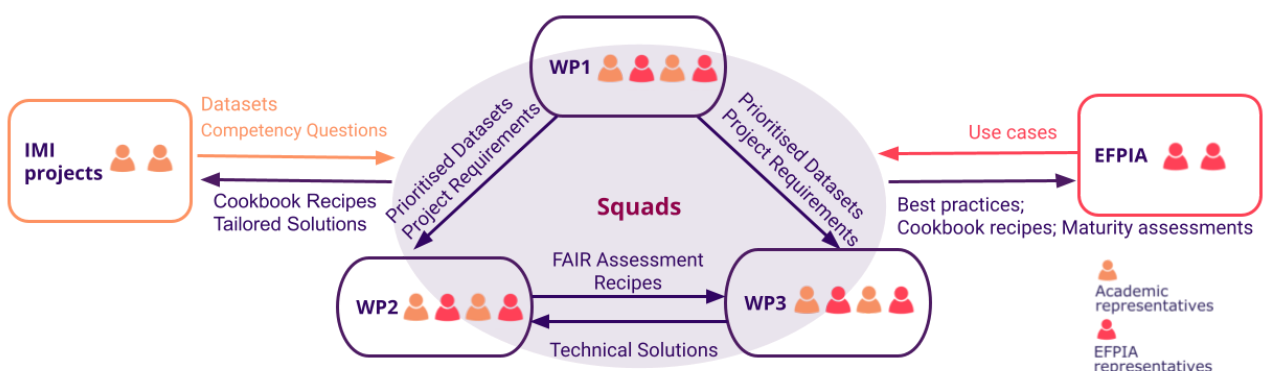
### 3. Overall Approach

The breadth of potential FAIRification activities FAIRplus could engage in, and the diversity of processes that are required across diverse IMI presents a challenge. Given this diversity, it was recognised from the point of project conception that FAIRplus would need to adopt an incremental approach to development, with frequent validation for technical feasibility, so as to ensure solutions could be effectively targeted. As such, the FAIRplus work package structure was designed to effectively “feed forward” development work through data selection, standards and processes, implementation and into outreach. At the same time, FAIRplus was designed to incorporate feedback loops from project outreach and dissemination activities into subsequent development activities of the technical and implementation work packages. The overall cyclic design of FAIRplus is shown in figure 1, and these cycles form part of an approach intended to ensure regular technical feasibility validation.



**Figure 1.** The FAIRplus development cycle, showing the flow of datasets through selection (WP1), standards and metrics planning (WP2), implementation (WP3) and dissemination (WP4)

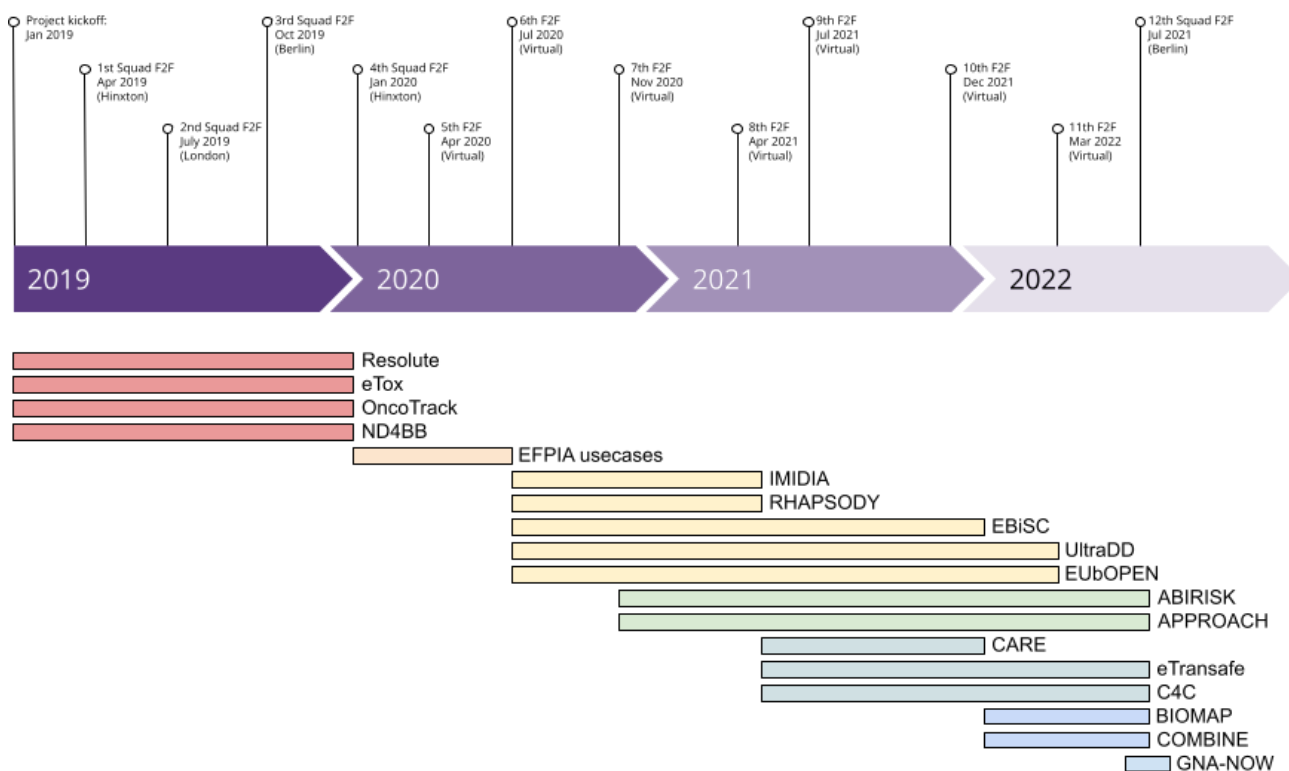
To ensure this cyclic development pattern could be made operational at a practical level, within FAIRplus we established a BYOD “squads” methodology (see D2.2). FAIRplus squads teams established several practical communication channels to facilitate the creation of feedback loops and ensure detailed insight into the feasibility of FAIRplus processes and technical developments from key stakeholders, including IMI and EFPIA partners. These feedback channels are shown in Figure 2.



**Figure 2.** The approach taken by FAIRplus to the continuous and incremental delivery of technical solutions, showing interaction and validation between the squad teams (drawn from WP1, 2 and 3 personnel) and IMI and EFPIA partners. This method of co-production method provides early and often feedback and validation to ensure solutions are technically feasible

The creation of squads has established an evolutionary approach to technical

implementation with constant, iterative validation of FAIRplus solutions. Further, to establish a regular cadence of working patterns and feedback, the squad teams created a 3 month long development cycle, with each cycle representing an area of focus for working teams and expected feedback. Areas of focus include FAIRification of specific IMI or EFPIA projects, as well as capacity building in one of the four impact areas (FAIRification processes, FAIR assessment, FAIR Cookbook and FAIR tooling and hosting solutions) defined above and in further detail in the sections below. Details on the squad cadence, along with squad workshops, are discussed further in D2.2 and the timeline of engagement and technical feasibility validation is shown in Figure 3, below.



**Figure 3.** Squad timelines, showing approximately 3-month release cycles marked by face to face meetings. The IMI and EFPIA projects that squads engaged with and used to validate technical solutions on are shown on the bottom of this timeline

## 4. Feasibility and Requirements

### FAIRification processes

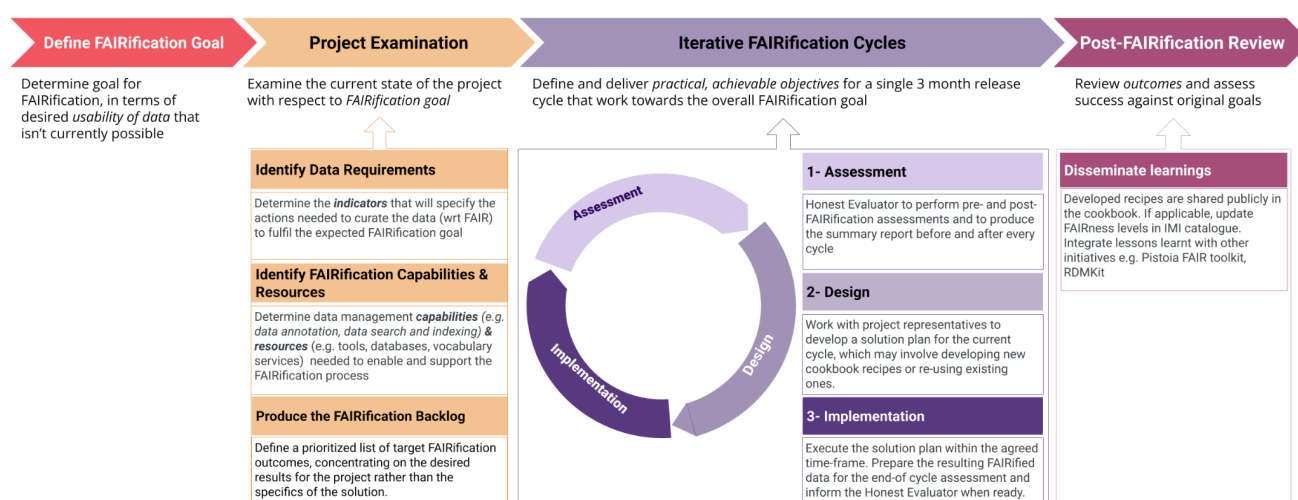
#### Impact

The core FAIRplus technical approach to FAIRification is encapsulated in the FAIRplus FAIRification process. The FAIRification process represents a shared framework through which projects and organisations can improve the FAIRness of their data assets. Having a shared, suitably generic FAIRification process that can be widely adopted ensures that organisations can share best practices and techniques, as well

as learn from lessons seen elsewhere. In this deliverable, we orientate the feasibility analysis of various technical implementations to the shared FAIRification process.

## Creation

It is most crucial that adopters of the FAIRplus approach can apply the FAIRplus FAIRification process within their organisations, and as such this is the “keystone” of technical feasibility within FAIRplus.



**Figure 4.** The current FAIRplus FAIRification process, showing two distinct phases; the “orange” preparation phase (identification of FAIRification goals and project examination) and the “purple” cyclic phase (design and implementation of incremental FAIRification solutions).

The FAIRplus FAIRification process (figure 4) now consists of two major phases, the “orange phase” and the “purple phase”. Orange phase is designed to establish and formalise long term FAIRification goals, and examine the current FAIR environment in which the project operates. The initial step of the orange phase, “*Define FAIRification goal*”, is designed to ensure FAIRification outcomes are tailored to the specific scientific or business objectives of the project or organisation. Then, the “*Project Examination*” step is designed to surface capabilities that are already in place within a project, and to collect information on which capabilities may be easiest to add, improve or adapt. Together, the orange phase steps are intended to produce a value- and cost- based assessment of FAIRification activities, helping subsequent process steps to demonstrate appropriate return on investment.

## Application

Application of the FAIRification process to IMI and EFPIA projects and data was performed on the 17 projects shown in table 1.

Project	Year (first engagement)	FAIRification Phase
Resolute	2019	Finished



eTox	2019	Finished
OncoTrack	2019	Finished
ND4BB	2019	Finished
IMIDIA	2020	Finished
RHAPSODY	2020	Finished
EBISC	2020	Finished
ABIRISK	2020	Finished
APPROACH	2020	Finished
UltraDD	2020	Purple
EUbOPEN	2020	Purple
CARE	2021	Finished
C4C	2021	Orange
COMBINE	2021	Purple
BIOMAP	2021	Purple
eTRANSafe	2021	Orange
GNA-NOW	2021	Purple

**Table 1.** Table showing the set of IMI projects with which FAIRplus has engaged to date, and to which the FAIRplus FAIRification process has been applied. Ongoing engagements are also listed ('Orange' and 'Purple'). Projects listed as "finished" have completed orange and purple phase, and are either in Post-FAIRification review or have case studies written up and disseminated (see D4.5)

Based on feedback from these projects, the FAIRification process has undergone several adaptations, resulting in a total of 6 major version releases. We previously reported on the FAIRification process in D3.1 (version 2.0) and D3.3 (version 3.0); the current version shown in figure 4 is version 6.3. The evolution of the FAIRification process is shown in Appendix A.

## Evaluation

The variety and number of projects that the FAIRplus FAIRification process has been applied to demonstrates its technical feasibility. Further, as part of the FAIRplus activities, we have seen that FAIRification processes are more feasible and more likely to be successful during initial data management planning (so called "prospective FAIRification") rather than after data has been generated ("retrospective FAIRification" or curation).

## Requirements

In order to adopt the FAIRplus process, projects require access to a small number of expert personnel of the “data steward” profile - these skills are necessary to correctly and successfully apply the FAIR data management techniques that are a prerequisite for the process.

## Summary of Feasibility Analysis

**Ease of adoption:** Easy

**Conditions for success:** Apply when data generation is ongoing or not yet started

**Prerequisites:** Experienced data stewards or data managers embedded within the projects

## FAIR assessment

### Impact

We outlined in D3.2 (FAIR metrics publication) that being able to perform an objective FAIR assessment can provide a metric that helps to evaluate and validate the FAIRification technical solutions. In D3.2, we also outlined our approach for conducting FAIR assessments using RDA indicators<sup>2</sup>. As we discussed in D3.3, it is critically important to be able to guide FAIRification initiatives and the production of technical solutions with an objective cost/benefit assessment, to avoid the risk of FAIRification becoming a “neverending story”.

### Creation

Techniques for FAIR assessment, using the FAIR Dataset Maturity (FAIR-DSM) model, shall be fully outlined in D2.6 (FAIR-CMM) in month 48. The FAIR-DSM model<sup>3</sup> (shown in Figure 5, below) is a reference model for state-of-FAIRness maturity improvement in research datasets. Classified by maturity levels, its associated indicators can be used to perform a FAIR assessment, with associated tooling<sup>4</sup> to make this process slightly simpler for a user.

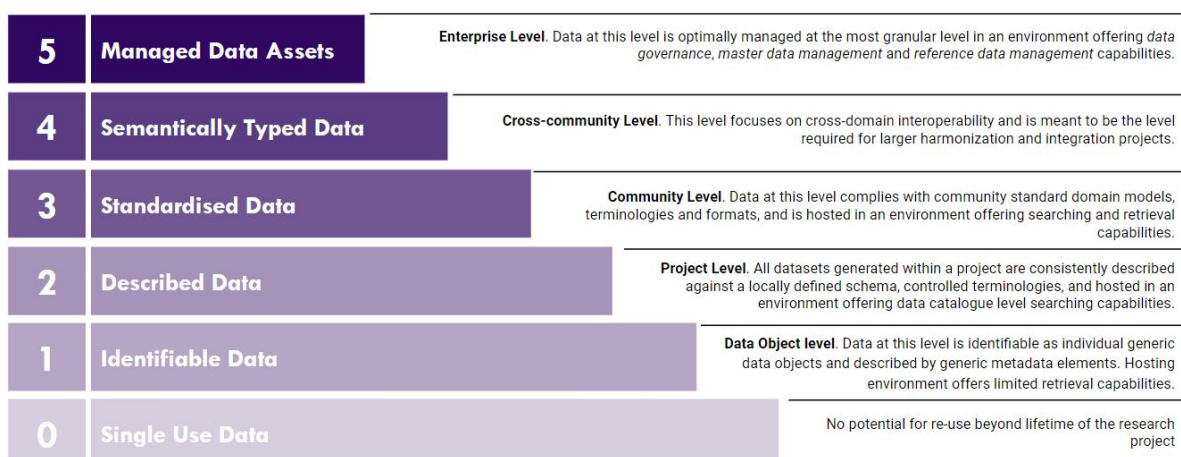
---

<sup>2</sup>

<https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines-0>

<sup>3</sup> <https://fairplus.github.io/Data-Maturity/>

<sup>4</sup> <https://github.com/FAIRplus/FAIR-DSM-Assessment-Tool>



**Figure 5.** The FAIR-DSM maturity model, showing 5 levels of maturity above the baseline state (level 0). Each maturity level has a set of associated indicators that can be used to perform a FAIR assessment.

## Application

Such FAIR assessments are now a routine part of the orange and purple phases of the FAIRification process (as shown in figure 4) and allow technical teams to continuously evaluate progress towards the maturity level mandated by the desired, formalised (FAIRification) goal, or else identify work necessary to reach the next maturity level. Table 2 shows a selection of projects that underwent a FAIR assessment using the latest FAIR-DSM version at a recent FAIRplus Squads meeting (December 2021).

Project Name	Initial Maturity Level	Post-FAIRification Maturity Level
ReSOLUTE	0	2
eTOX	0	2
IMIDIA	1	1
RHAPSODY	1	1
ABIRISK	0	0
CARE	0	3
EBiSC	0	1

**Table 2.** A selection of projects which have undergone a recent FAIR-DSM maturity assessment pre- and post-FAIRification, showing the change in maturity level that occurred as a result of FAIRification work and demonstrating the feasibility of applying such assessments.

## Evaluation

FAIR assessments are vital to guide design and implementation work and the production of technical solutions within the FAIRification process. However,

assessments are also used in the orange phase to surface impactful opportunities for investment. From our experience with 17 IMI projects so far, discussions in the orange phase often reveal that whilst there is an organisational mandate to increase FAIR maturity, there is often no specific goal being targeted. In these scenarios, running an exploratory FAIR assessment can identify quick wins that can be prioritised for development to rapidly promote datasets to higher maturity levels. Assessments have also been used to guide and prioritise developments. D3.2 discussed the use of metrics and FAIR assessments in this manner, including discussion of the challenges that can be encountered. Table 2 shows that some projects, despite improving numerous metrics after FAIRification work, did not increase in overall maturity. However, the fact that this discrepancy was highlighted through pre- and post-assessments is further validation of the technical feasibility of this approach, even if different choices about the work that was done, possibly resulting in a greater maturity increase, could have been made.

## Requirements

Running a FAIR-DSM assessment is an expert activity. From working with a number of IMI projects and validating early versions of the FAIR-DSM, we have seen that assessments can be difficult for inexperienced individuals to run. The FAIR-DSM model and its assessment tool assume a level of familiarity and experience with data curation concepts around data and metadata that require prior expertise. In D3.2, we stated that “Performing an automated assessment, in addition to manual review and further use-case based evaluation of the dataset FAIRness will provide the most scalable result.” Such automated assessments are still a laudable goal, but not yet fully within reach, and until they are, semi-automated and manual assessments require individuals to reason about their data and metadata in an abstract way; not everyone has the ability to do so.

## Summary of Feasibility Analysis

**Ease of adoption:** Medium

**Conditions for success:** Prior experience with running FAIR assessments and data audits. Clearly delineated datasets for assessment.

**Prerequisites:** Experienced data stewards who can learn the FAIR-DSM assessment process and understand the definitions of terms used in the assessment. Access to FAIRplus personnel to explain the current assessment model

## FAIR Cookbook

### Impact

The FAIR Cookbook is described in detail in D2.1, and aims to provide guidance to researchers and data stewards of life science data, going beyond the aspirational and general FAIR principles into specific, actionable recipes that can be applied to datasets to increase their FAIRness. This type of specific guidance forms a vital part of the

FAIRification process - recommendations on design and implementation in the purple phase of the FAIRplus process are critical to support FAIR best practices and the sharing of FAIR data management solutions.

## Creation

The technical framework for the FAIR Cookbook<sup>5</sup> was established at the start of the project, and a small number of recipes added to capture the experiences of FAIRifying the FAIRplus pilot projects. Since then, over 70 recipes have been added incrementally, capturing experiences of FAIRifying further IMI datasets, the experience of EFPIA partners, the fellowship programme and more. Since March 2021, FAIRplus Cookbook Book Dash events have been run approximately monthly to expand the cookbook, create new recipes, and refine existing recipes in response to feedback from implementers. These events have also targeted the user experience and improved metadata, facilitating cookbook search.

## Application

More than 70 recipes exist in the cookbook, each with its own defined audience and expected outcomes. Many of these recipes have been applied to IMI projects by the FAIRplus squads. Table 3 shows a list of IMI projects FAIRplus has worked with, and the recipes that have been applied to FAIRify datasets in each instance.

Project Name	Recipes Applied / In Preparation
Resolute	JSON schemas for transcriptomics datasets ( <a href="#">FCB045</a> <sup>6</sup> ); Requesting new ontology terms( <a href="#">FCB021</a> <sup>7</sup> ); Safe data transferring and accessing ( <a href="#">FCB014</a> <sup>8</sup> )
eTox	Chemical validation ( <a href="#">FCB007</a> <sup>9</sup> ); Omics datasets FAIRification experience( <a href="#">FCB042</a> <sup>10</sup> )
OncoTrack	Metadata extraction and deposition ( <a href="#">FCB044</a> <sup>11</sup> ); Cohort metadata ( <a href="#">FCB044</a> <sup>11</sup> )
ND4BB	Chemical activity dataset ETL ( <a href="#">FCB043</a> <sup>12</sup> )
IMIDIA	Ontology selection ( <a href="#">FCB019</a> <sup>13</sup> )

<sup>5</sup> <https://faircookbook.elixir-europe.org/>

<sup>6</sup> <http://w3id.org/faircookbook/FCB045>

<sup>7</sup> <https://w3id.org/faircookbook/FCB021>

<sup>8</sup> <https://w3id.org/faircookbook/FCB014>

<sup>9</sup> <https://w3id.org/faircookbook/FCB007>

<sup>10</sup> <https://faircookbook.elixir-europe.org/content/recipes/applied-examples/etox.html>

<sup>11</sup> <http://w3id.org/faircookbook/FCB044>

<sup>12</sup> <http://w3id.org/faircookbook/FCB043>

<sup>13</sup> <https://w3id.org/faircookbook/FCB019>

RHAPSODY	Ontology selection ( <a href="#">FCB019</a> <sup>14</sup> )
EBISC	Dataset markup with Bioschemas ( <a href="#">FCB011</a> <sup>15</sup> ); Search engine optimisation ( <a href="#">FCB010</a> <sup>16</sup> ); Ontology recommendation( <a href="#">FCB020</a> <sup>17</sup> );
ABIRISK	Data dictionary ( <a href="#">FCB025</a> <sup>18</sup> ), Data licensing ( <a href="#">FCB034</a> <sup>19</sup> ), Data use conditions ( <a href="#">FCB035</a> <sup>20</sup> )
APPROACH	Data dictionary ( <a href="#">FCB025</a> <sup>18</sup> ), Data licensing ( <a href="#">FCB034</a> <sup>19</sup> ), Data use conditions ( <a href="#">FCB035</a> <sup>20</sup> ), CDISC ETL recipe (in preparation)
UltraDD	DMP guidance (advice includes <a href="#">recipe</a> <sup>21</sup> )
EUbOPEN	Image data submission ( <a href="#">FCB067</a> <sup>22</sup> )
CARE	BioActivity data profile ( <a href="#">FCB057</a> <sup>23</sup> )
C4C	Data dictionary ( <a href="#">FCB025</a> <sup>18</sup> ); Clinical trial data type recipe (in preparation)
COMBINE	Ontology selection ( <a href="#">FCB019</a> <sup>14</sup> ); Application ontology ( <a href="#">FCB023</a> <sup>24</sup> )

**Table 3.** List of IMI projects to which recipes have been applied, and the recipes that were used during FAIRification work

## Evaluation

Cookbook recipes have been continuously adapted and revised based on feedback from IMI and EFPIA projects, and participants in the fellowship programme (see D4.3). Recipes themselves have therefore been shown to be technically feasible to apply, although there is variability in the ease with which recipes can be applied. This difficulty is also captured in the recipes themselves (see also Figure 6).

<sup>14</sup> <https://w3id.org/faircookbook/FCB019>

<sup>15</sup> <https://w3id.org/faircookbook/FCB011>

<sup>16</sup> <https://w3id.org/faircookbook/FCB010>

<sup>17</sup> <https://w3id.org/faircookbook/FCB020>

<sup>18</sup> <https://w3id.org/faircookbook/FCB025>

<sup>19</sup> <https://w3id.org/faircookbook/FCB034>

<sup>20</sup> <https://w3id.org/faircookbook/FCB035>

<sup>21</sup>


<https://faircookbook.elixir-europe.org/content/recipes/infrastructure/id-resolution.html?highlight=identifiers>


<sup>22</sup> <http://w3id.org/faircookbook/FCB067>



<sup>23</sup> <https://w3id.org/faircookbook/FCB057>

<sup>24</sup> <https://w3id.org/faircookbook/FCB023>

Recipe Overview


 Reading Time  
60 minutes


 Executable Code  
Yes


 Difficulty  


## Building an application ontology with

### ROBOT

 Recipe Type  
Hands-on

 Audience  
Terminology Manager, Data Manager, Data Scientist, Ontologist

 Maturity Level & Indicator  
[F+MM-1.1C] [F+MM-1.2C]

Cite me with [FCB023](#)

**Figure 6.** A header from an example recipe in the FAIRplus Cookbook, showing the audience for this recipe and the difficulty of this recipe to apply in practice (4 out of 5).

## Requirements

In order to adopt recommendations from FAIR Cookbook recipes, IMI projects, EFPIA partners or other organisations will typically need dedicated data management personnel of the right persona (e.g. data manager, ontologist) as listed in the recipes. It is common in most projects to have personnel devoted to data management and a formal data management plan; as long as projects are suitably resourced for data management, this is not a challenge. Note that projects which are seeking to exploit data that has been generated in the past often will not gain maximum benefit from cookbook recipes, which teach good data management practices rather than curation techniques. More advanced recipes require deep research data management expertise and sometimes specific technical skills.

## Summary of Feasibility Analysis

**Ease of adoption:** Medium

**Conditions for success:** Active data generation projects (i.e. not a cohort of archive data that needs to be curated). An environment in which data managers are empowered to make data management changes

**Prerequisites:** Experienced FAIR practitioners in data management/data stewardship roles, with the expertise to apply recipe techniques (including sometimes advanced techniques)

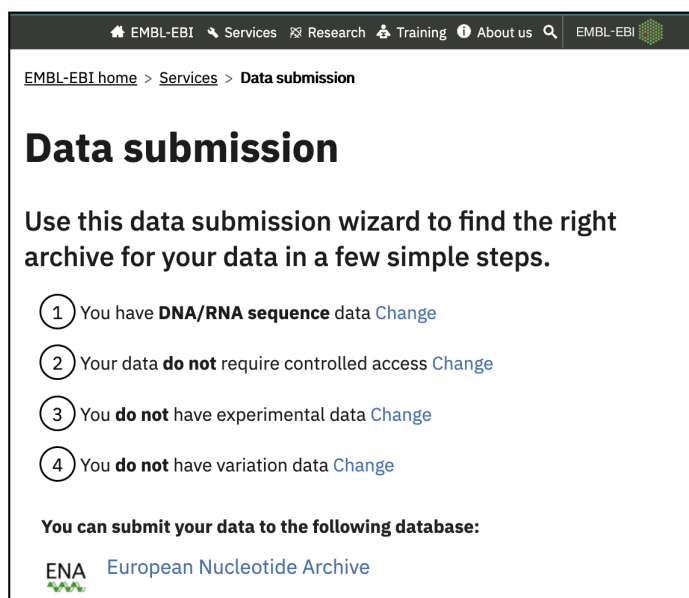
## FAIR tooling and data hosting

### Impact

Where data is hosted and how that data is made accessible has a major impact on its FAIRness. Many projects neglect these elements; instead of modelling data in a rich database or knowledge graph, the data is often locked away inside a project-specific database that becomes unsustainable once the project ends. For this reason, the *“Hosting Environment Capabilities”* category is a first class citizen in the FAIR-DSM maturity model. The FAIRplus FAIRification process emphasises examination of the project capabilities to determine hosting requirements and encourages deposition of data into a sustainable FAIR data repository as early as possible so as to ensure longevity and sustainability of produced data assets. Combined with hosting solutions, there are also a wide variety of FAIR data management tools available to support data FAIRification, harmonisation and submission to public repositories. Such tools make FAIRification tasks more accessible to data managers and research software engineers, but these tools are not always well known.

### Creation

To facilitate the identification of good hosting solutions, the FAIR-DSM model and the FAIR Wizard (see D3.7) both emphasise the importance of sustainable data hosting. Resources like the EMBL-EBI submissions service<sup>25</sup> also help identify a long term hosting solution by datatype.



**Figure 7.** The EMBL-EBI submissions wizard, to identify a long-term sustainable hosting solution by datatype (in this case, non-controlled access sequence data can be submitted to ENA)

<sup>25</sup> <https://www.ebi.ac.uk/submission/>



Along with data hosted solutions, within FAIRplus we have also developed a new FAIR tool discoverer<sup>26</sup>, providing search for tools that support typical FAIRification activities. This tool provides a search interface over common bioinformatics tools registries (e.g. bio.tools<sup>27</sup>) and supports queries by typical FAIRification tasks. This complements the tool descriptions included in FAIR Cookbook recipes, which explain which tools to use to fulfil certain tasks.

## Application

Hosting solutions have been identified for 6 datasets from IMI projects that have engaged with the FAIRplus project, as shown in Table 4.

Project Name	Datatype	Access Model	Hosting solution
Resolute	Transcriptomics, proteomics, metabolomics	public	BioProjects, BioSamples and SRA (Project: <a href="https://www.ncbi.nlm.nih.gov/bioproject/545487">PRJNA545487</a> <sup>28</sup> )
eTox	Chemical compounds, toxicology assays	public (sampler), managed	<a href="http://etoxsys.eu/">Self-hosted http://etoxsys.eu/</a>
OncoTrack	Oncology (Transcriptomics, proteomics, metabolomics)	managed	EGA (Study: <a href="https://ega-archive.org/studies/EGAS00001001752">EGAS00001001752</a> <sup>29</sup> )
ND4BB	Chemical compounds	public	Self-hosted <a href="https://www.dsf.unica.it/translocation/db/">https://www.dsf.unica.it/translocation/db/</a>
IMIDIA	Clinical data, transcriptomics	managed	n/a
RHAPSODY	Clinical data, transcriptomics	managed	n/a
EBISC	Cell line metadata, genomics	public	<a href="https://ebisc.org/">Self-hosted https://ebisc.org/</a>
ABIRISK	Clinical data, biomarkers, transcriptomics	managed	n/a
APPROACH	Clinical data, imaging data, biomarkers	managed	n/a
UltraDD	Chemical compounds	managed	n/a

<sup>26</sup> <https://fair-tool-discoverer.bsc.es/>

<sup>27</sup> <https://bio.tools/>

<sup>28</sup> <https://www.ncbi.nlm.nih.gov/bioproject/545487>

<sup>29</sup> <https://ega-archive.org/studies/EGAS00001001752>

EUbOPEN	Chemical compounds	managed	n/a
CARE	Bioassay data, chemical compounds	public	ChEMBL (Report card: <a href="#">CHEMBL4651402</a> <sup>30</sup> )
C4C	Clinical data	managed	n/a
COMBINE	Microbial resistance bioassays Ontology	managed	n/a
BIOMAP	Clinical data, transcriptomics	managed	n/a
eTRANSafe	Toxicity, transcriptomics	managed	n/a
GNA-NOW	Microbial resistance data	managed	n/a

**Table 4.** A list of IMI projects along with their datatypes and the hosting solution that was identified for datasets from these projects, including links to those datasets once brokered

In each case, data has been FAIRified and brokered to the named hosting solution.

Tools have been utilised when mandated by recipes such as those shown in table 4 above. Notably, the ROBOT tool<sup>31</sup> was used effectively in EFPIA as described in D3.4. The tool discoverer is a new application and is currently undergoing user experience testing to determine its effectiveness when deployed as part of FAIRification activities.

## Evaluation

The correct identification of data hosting solutions requires some data management expertise, but there are many good guidelines and services designed to support this activity. Most data managers will be highly experienced in submitting data to public archives or repositories. The greater challenge is in encouraging or persuading them to do so, rather than developing a new bespoke project-internal database to manage data assets: whilst these databases are frequently needed during the life cycle of the project, data assets are much more sustainable when data is captured using standards that are compatible with those of public repositories, and data is brokered into public repositories before the project ends.

A major challenge is ensuring the open availability of data from human subjects, which is classified as managed data and often difficult to share broadly under GDPR. Whilst there are managed access repositories such as the European Genome-Phenome Archive, EGA<sup>32</sup>, for these data, concerns over the regulatory environment, differing national interpretations of GDPR and the international nature of IMI projects often create barriers to prompt release of data. This is clearly illustrated in table 4; many projects with managed access data have no identified hosting solution. This has also presented a technical challenge for FAIRplus;

<sup>30</sup> [https://www.ebi.ac.uk/chembl/document\\_report\\_card/CHEMBL4651402/](https://www.ebi.ac.uk/chembl/document_report_card/CHEMBL4651402/)

<sup>31</sup> <http://robot.obolibrary.org/>

<sup>32</sup> <https://ega-archive.org/>

FAIRification practices have often been designed “in principle” to support the FAIRification of unseen data. However, where hosting solutions can be identified, it is simple to apply them.

## Requirements

Suitable hosting solutions and tool usage have no requirements other than personnel to perform data submissions. Where FAIRification tools can be applied, these should be aligned with the application of suitable recipes as defined in the cookbook section above.

## Summary of Feasibility Analysis

**Ease of adoption:** Easy

**Conditions for success:** Processes in place for timely submission of data to public archives

**Prerequisites:** None

## 5. Discussion

We have demonstrated that the FAIRplus FAIRification process can be applied to a large variety of projects successfully. Applying the FAIRplus FAIRification process requires several technical implementation activities; namely, the use of the FAIR-DSM maturity model (D2.6), the use of recipes from the FAIR Cookbook (D2.1) and the use of hosting solutions and FAIR tools, including the IMI Data Catalog (D3.5).

Our co-production model is designed to provide early, continuous feedback on the technical feasibility of the solutions we have produced. This method, further described in D2.2, optimises the production of solutions that are feasible and prunes out infeasible approaches very early on.

Through our engagement with a variety of IMI projects, we have seen that projects tend to begin a FAIRification journey with either

- a) a clear statement of their expected business objectives (goal-driven) or
- b) a general strategic aim to improve their overall FAIRness (data-driven)

Goal-driven projects need early guidance about the capabilities they need to focus on, and can utilise guidance tools (see D3.7) to support them. These projects tend to rely on good recipes, tools and hosting solutions to apply the FAIRification process, and (as long as data generation is ongoing) have shown good success employing these solutions.

Data-driven projects need suggestions on which capabilities they should improve to increase their level of FAIRness from its current state, and rely on the FAIR assessments using the FAIR-DSM model to accomplish this. Projects adopting the FAIR-DSM and using it to define “quick wins” have demonstrated that this is highly

feasible and have shown very good success - CARE is a notable example, where the FAIR-DSM was used to guide simple changes. Adopting the ChEMBL data standard and brokering datasets to the ChEMBL database caused CARE datasets to rise from a FAIR-DSM maturity level 0 to level 3 with minimum effort (see table 2).

Prompt release of managed access data remains a major issue. This is a community-wide issue that has been well documented<sup>33</sup>, but nonetheless presents issues for FAIRification: it is difficult to undertake FAIRification processes without being able to evaluate progress through examination of the data before and after. Databases like EGA represent part of the solution; but smoother legal processes around data sharing and access are required to be able to truly apply the benefits of standardised FAIRplus technical solutions to managed access data.

## 6. Conclusion

Technical solutions produced by FAIRplus fall into one of four categories:

- 1. FAIRification processes**
- 2. FAIR assessments**
- 3. FAIR Cookbook recipes**
- 4. FAIR tooling and hosting solutions**

The co-production model utilised by FAIRplus, through the BYOD methodology and use of “squad teams”, has ensured that investment is targeted towards only technically feasible solutions, and infeasible solutions do not survive through multiple iterations of the development process. We have demonstrated that, in each of these four impact areas, FAIRplus has produced technically feasible solutions that fit together into a generalised FAIRification framework that can be applied at scale in a wide variety of environments, projects and organisations.

Whilst we have demonstrated that FAIRplus solutions are clearly technically feasible and can be applied by both IMI and EFPIA, we have also seen that the ease with which these solutions can be applied varies, and in some cases requires a high level of expertise or prior experience, for example when applying advanced cookbook recipes or the FAIR-DSM model to assessments. Making the FAIRplus FAIRification framework accessible is a clear remaining area of opportunity. In D3.7, we report on the FAIR Wizard, a FAIRification guidance tool designed to lower the barrier of entry to applying the FAIRplus FAIRification framework that binds a number of tools together, and we will continue to develop this tool in the remainder of the project.

---

<sup>33</sup> “The broken promise that undermines human genome research”, Kendall Powell, Nature, 2021  
<https://doi.org/10.1038/d41586-021-00331-5>

# Appendix A - Evolution of the FAIRplus FAIRification Process

The illustrations below show the evolution of the FAIRplus FAIRification process. Some versions have been used more extensively (notably v2.0, v3.0 and v6.3) and hence have more polished illustrations, whereas others have been used mostly for internal alignment and management of activities (e.g. v1.0). The evolution of this process demonstrates the utility of the mechanism by which FAIRplus has incrementally improved and adapted its processes, through feedback via the squads framework, to ensure technical feasibility.

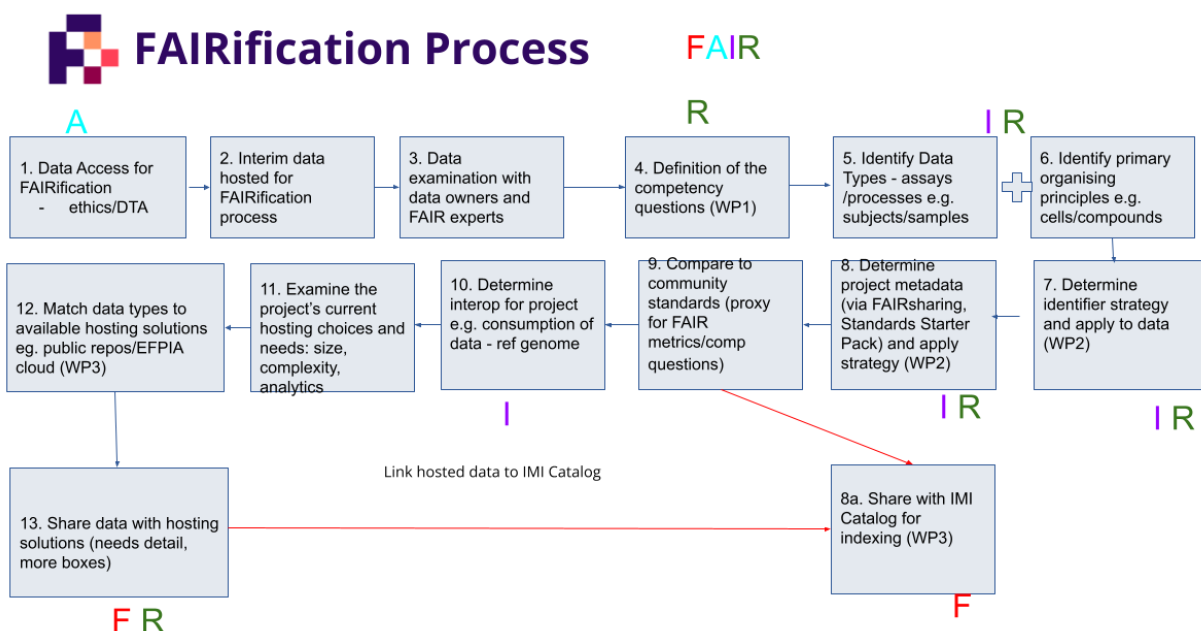


Figure A1: FAIRplus FAIRification process v1.0

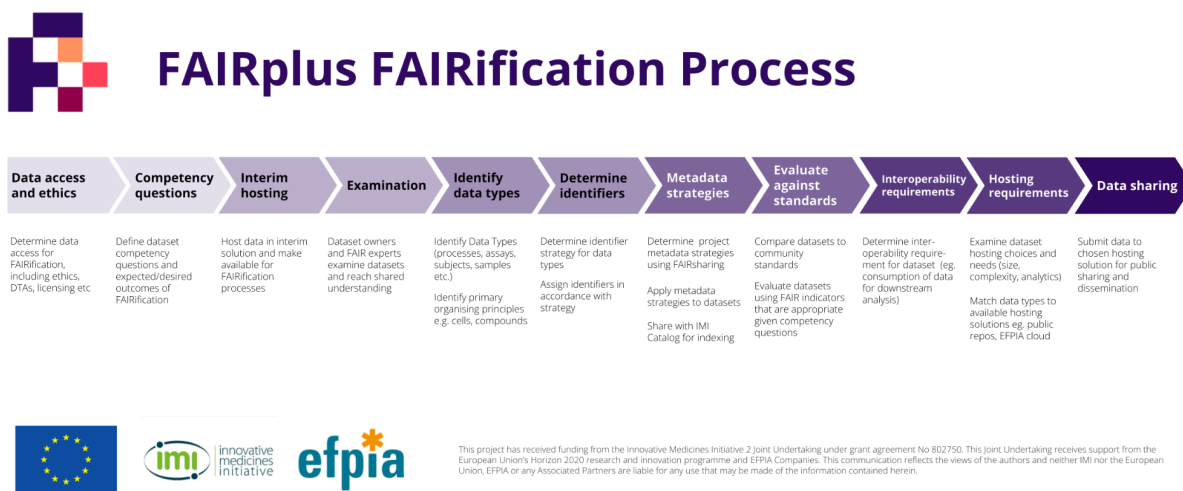


Figure A2: FAIRplus FAIRification process v2.0, included in D3.1



Figure A3: FAIRplus FAIRification process v3.0 (included in D3.3)

## FAIRplus FAIRification Process

Changes from v3.0



Figure A4: FAIRplus FAIRification process v4.0

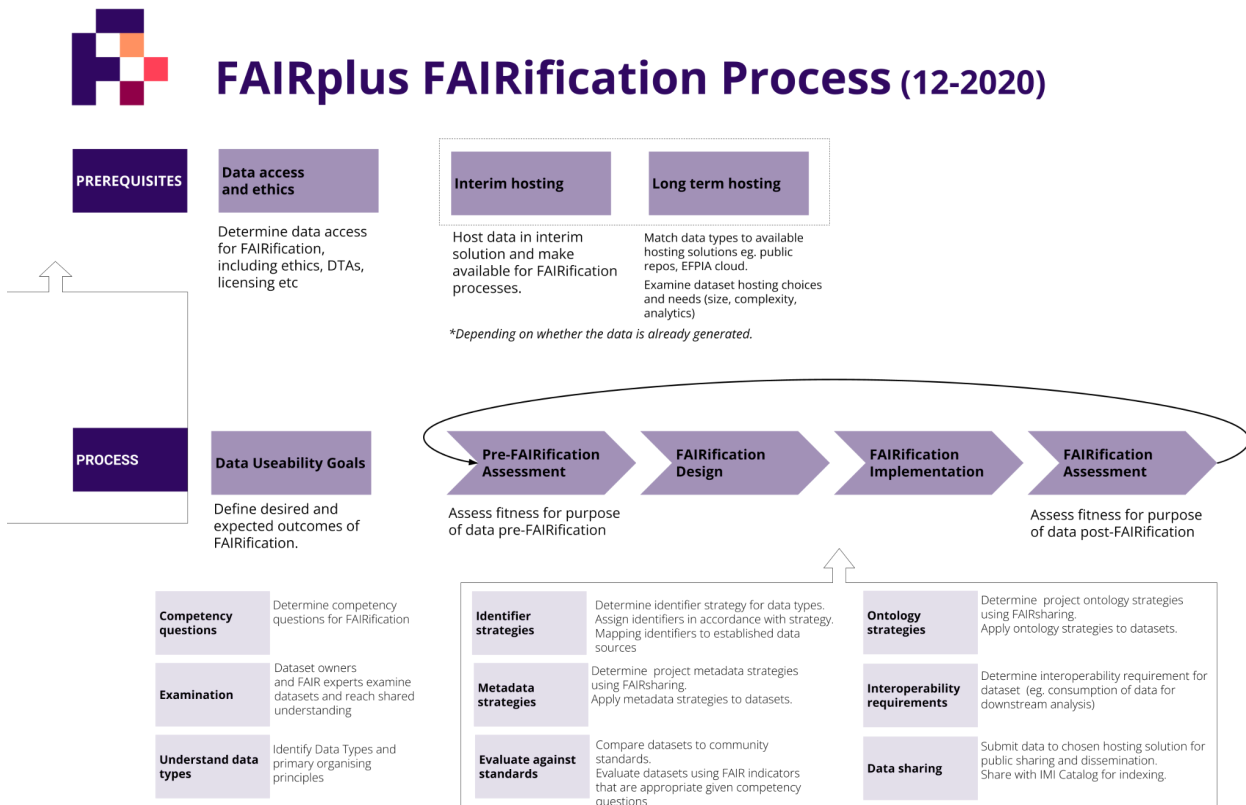


Figure A5: FAIRplus FAIRification process v5.0

Figure A6: FAIRplus FAIRification process v6.3 (the current version, included in D3.4)