IMI2 Project 802750 - FAIRplus

FAIRification of IMI and EFPIA data

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

D3.1 First phase exemplar IMI projects FAIRified

Lead contributor	Tony Burdett (1 – EMBL (EBI)
	tburdett@ebi.ac.uk
Other contributors	Mélanie Courtot (1 – EMBL (EBI)
	Fuqi Xu (1 – EMBL (EBI)
	Wei Gu (7 – University of Luxembourg)
	Venkata Satagopam (7 – University of Luxembourg)
	Nick Juty (4 - University of Manchester)
	Oya Beyan (2 - Fraunhofer)
	Francesco Ronzano (9 - IMIM)

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No. 802750. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation and EFPIA companies.

Laura Furlong (9 - IMIM)
Ibrahim Emam (11 - Imperial College London - Data Science Institute)
Carole Goble (4 - University of Manchester)
Robert Giessmann (21 - Bayer)
Nick Lynch (10 - Open PHACTS Foundation)

Due date	31 Dec 2019
Delivery date	20 Dec 2019
Deliverable type	R
Dissemination level	PU

Description of Work	Version	Date
	V1.0	19 Dec 2019

Document History

Version	Date	Description
V0.1	29 Nov 2019	First Draft
V0.2	10 Dec 2019	Initial comments addressed and final hyperlinks added
V0.3	13 Dec 2019	Final Draft
V1.0	19 Dec 2019	Final Version

Table of Contents

Document History 2

Executive Summary 3

Methods 4

Overview of Process 4

Recipes 6

eTOX 6

ND4BB 6

ReSOLUTE 6

OncoTrack 6

FAIR Assessments 6
eTOX 6
ND4BB 6
ReSOLUTE 6
OncoTrack 7

Results 7
Data Available 7
The results of the FAIR assessment of the four pilot projects have been published in the IMI data catalog and highlighted with a "FAIRplus-Evaluated" badge. FAIR scores (overall, mandatory and recommended indicators), FAIR indicators (version and link) and dataset links (when available) are included alongside project metadata, as shown in Figure 2. 7
eTOX 7
ND4BB 7
ReSOLUTE 7
OncoTrack 7

Discussion 8
Data Access Restrictions 8
Resulting Levels of FAIR 9
CMMI Evaluation 9

Conclusion 9

1. Executive Summary

FAIRplus seeks to establish 'FAIRification' processes that can be used at scale to ensure FAIRness of IMI data. In order to establish, refine and validate FAIRplus FAIRification techniques, four pilot IMI datasets were selected in D1.1[1]. This deliverable describes the outcomes achieved to date when FAIRifying these datasets using the newly established FAIRplus FAIRification process.

The FAIRified pilot datasets are now listed in the IMI data catalogue, along with an evaluation of their level of FAIR after FAIRification processes were applied. In this deliverable, we show that FAIRplus FAIRification processes generally increase the level of FAIR for pilot datasets, although no datasets became "completely FAIR" as a result. Level of FAIR is quantified by a series of FAIR indicators, established from a FAIR assessment conducted on each dataset before and after FAIRification. The results of these assessments are linked from the IMI data catalogue. All processes used during FAIRification are documented as recipes in the FAIR cookbook.

[1] https://zenodo.org/record/3274230#.XfvayIP7Q6h

2. Methods

Overview of Process

The four selected pilot projects (see deliverable D1.1) were each assessed, prior to FAIRification, for their level of FAIR using RDA FAIR indicators[2].

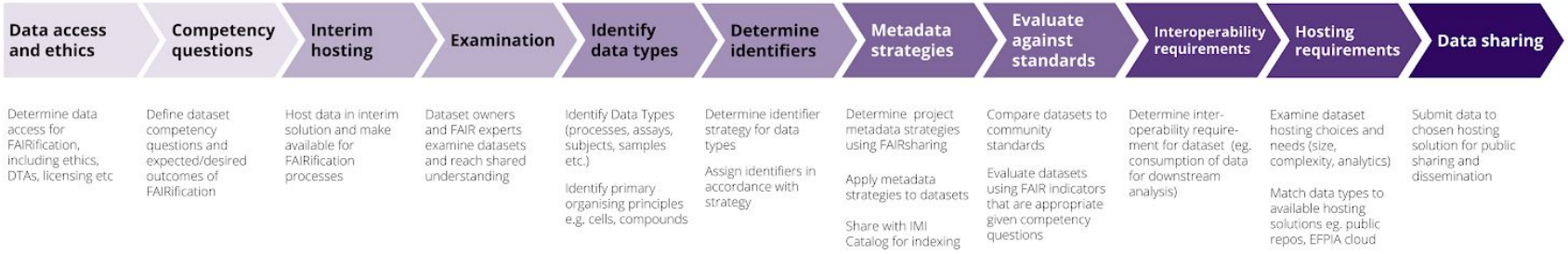
FAIRification recipes were produced for each selected project. These four pilot IMI projects satisfy a variety of eligibility criteria (the methodology is described in D1.1), and consequently each dataset has a different data profile, requiring a different set of recipes. Recipes were defined during FAIRification activities for these pilot projects; it is expected that they will be further refined as they are applied to future selected projects. Each recipe provides a dataset-specific implementation of the FAIRification process illustrated in Figure 1. As part of task 2.1.4[3], recipes produced will be continuously assessed and refined against the FAIRplus Capability Maturity Model (which will be produced as described in task 2.2.2).

Once FAIRification recipes had been applied to the four pilot projects, those projects were reassessed for their improved level of FAIR, using the same RDA FAIR indicators (or an updated version of those indicators).

The four pilot projects were made available from the IMI data catalog[4]. The catalog was updated to illustrate the level of FAIR (after FAIRification) for each project, and to provide links to the FAIR indicators used to formulate this judgement.

[2] https://github.com/RDA-FAIR/FAIR-data-maturity-model-WG
[3] https://fairplus-project.eu/about/how-project-organised#wp2
[4] https://datacatalog.elixir-luxembourg.org/

FAIRplus FAIRification Process



This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 802750. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA Companies. This communication reflects the views of the authors and neither IMI nor the European Union, EFPIA or any Associated Partners are liable for any use that may be made of the information contained herein.

Figure 1. The FAIRplus FAIRification process, showing a series of sequential steps to follow in order to improve the overall level of FAIR of an individual project, and defining the required outcomes for each step

Recipes

eTOX

FAIRification Recipe v1.0[5] for the eTOX dataset.

ND4BB

FAIRification Recipe v1.0[6] for the ND4BB AMR dataset.

ReSOLUTE

FAIRification Recipe_v1.0[7] for the ReSOLUTE dataset.

OncoTrack

FAIRification Recipe_v1.0[8] for the OncoTrack dataset.

FAIR Assessments

eTOX

Initial FAIR assessment results[9] for eTOX.
FAIR assessment results[10] for eTOX after FAIRplus FAIRification processes.

ND4BB

Initial FAIR assessment results[11] for ND4BB.
FAIR assessment results[12] for ND4BB after FAIRplus FAIRification processes.

ReSOLUTE

Initial FAIR assessment results[13] for ReSOLUTE.
FAIR assessment results[14] for ReSOLUTE after FAIRplus FAIRification processes.

[5] https://fairplus.github.io/the-fair-cookbook/recipes/etox_raw/FAIRplus_Recipe_Ontology_mapping_the_eTox_dataset_scenario
[6] https://fairplus.github.io/the-fair-cookbook/recipes/nd4bb_raw/FAIRification_CookBook_Recipe1_V02
[7] https://fairplus.github.io/the-fair-cookbook/recipes/resolute_raw/rawRecipe_resolute
[8] https://fairplus.github.io/the-fair-cookbook/recipes/oncotrack_raw/rawRecipe-oncotrack
[9] https://fairplus.github.io/fairification-results/2019-12-17-eTOX-pre-assessment/
[10] https://fairplus.github.io/fairification-results/2019-12-17-eTOX-post-assessment/
[11] https://fairplus.github.io/fairification-results/2019-12-17-ND4BB-Pre-assessment/
[12] https://fairplus.github.io/fairification-results/2019-12-17-ND4BB-post-assessment/
[13] https://fairplus.github.io/fairification-results/2019-12-17-RESOLUTE-Pre-assessment/
[14] https://fairplus.github.io/fairification-results/2019-12-17-RESOLUTE-post-assessment/

OncoTrack

Initial FAIR assessment results[15] for OncoTrack.
FAIR assessment results[16] for OncoTrack after FAIRplus FAIRification processes.

3. Results

Data Available

The results of the FAIR assessment of the four pilot projects have been published in the IMI data catalog and highlighted with a "FAIRplus-Evaluated" badge. FAIR scores (overall, mandatory and recommended indicators), FAIR indicators (version and link) and dataset links (when available) are included alongside project metadata, as shown in Figure 2.

eTOX

eTOX entry[17] in IMI Data Catalog.

ND4BB

ND4BB entry[18] in IMI Data Catalog.

ReSOLUTE

ReSOLUTE entry[19] in IMI Data Catalog.

OncoTrack

OncoTrack entry[20] in IMI Data Catalog.

[15] https://fairplus.github.io/fairification-results/2019-12-17-OncoTrack-Pre-assessment/
[16] https://fairplus.github.io/fairification-results/2019-12-17-OncoTrack-post-assessment/
[17] https://datacatalog.elixir-luxembourg.org/e/dataset/484548a2-1ce8-11ea-8039-8c8590c45a21
[18] https://datacatalog.elixir-luxembourg.org/e/dataset/79d1c3a2-104d-11ea-9e31-0050569a29db
[19] https://datacatalog.elixir-luxembourg.org/e/dataset/79d2691a-104d-11ea-9e31-0050569a29db
[20] https://datacatalog.elixir-luxembourg.org/e/dataset/64f33e4f-0d6d-4062-86c5-9c3db4e3a99a

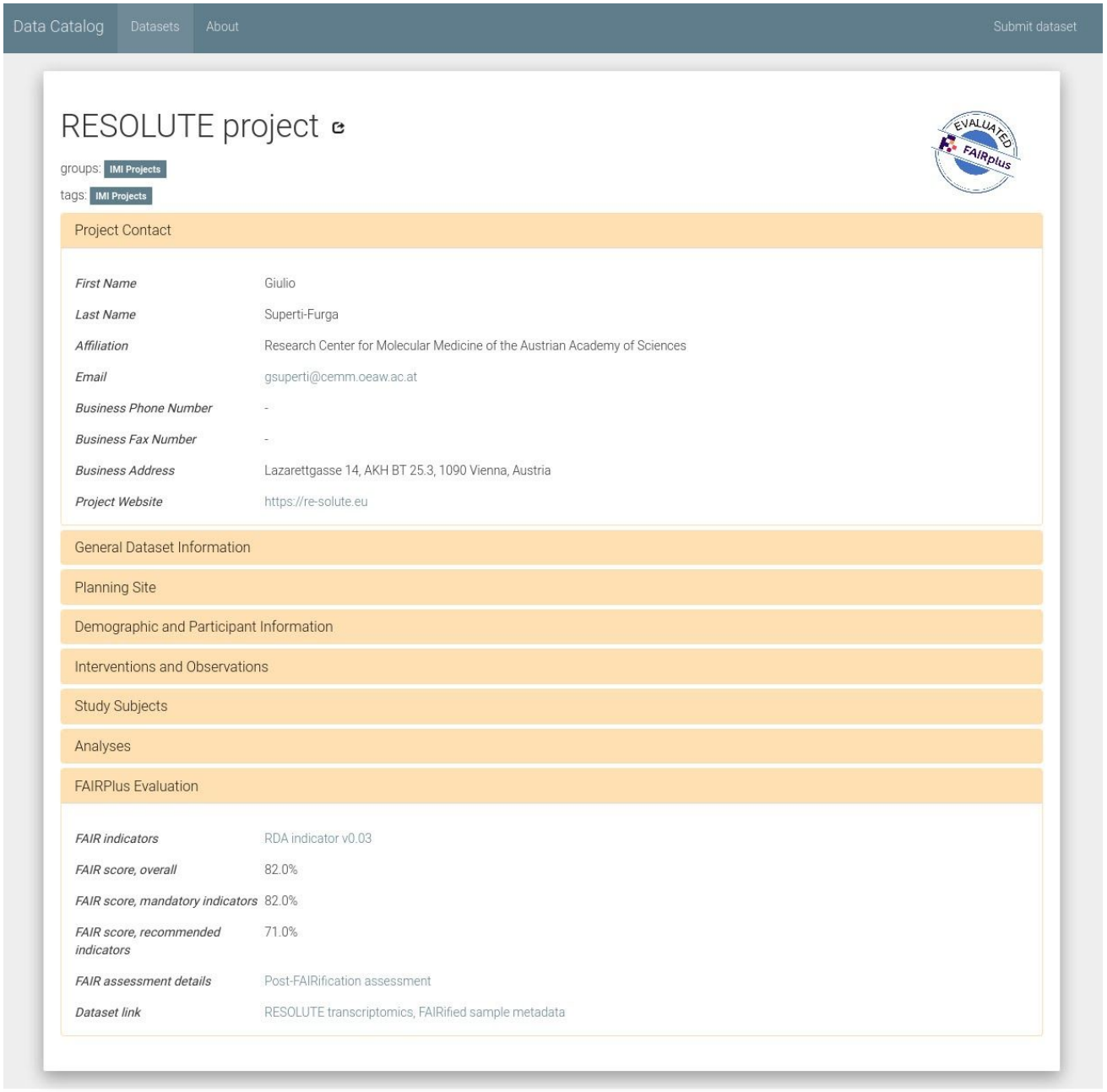


Figure 2. The ReSOLUTE project entry in the IMI data catalogue, showing the new "FAIRplus evaluated" badge and FAIRplus evaluation results, highlighting FAIR scores (assessed using RDA indicators v0.03) and dataset link alongside other project-level metadata.

4. Discussion

Data Access Restrictions

Legal concerns over the release of a dataset from the OncoTrack project meant that FAIR assessment and FAIRification processes were significantly delayed. Whilst a publicly available subset of OncoTrack data has been FAIRified, much less time was available to develop these recipes and the FAIRification processes for this dataset. We therefore expect that CMMI assessments will yield a lower level of maturity for

OncoTrack than for other projects.

Resulting Levels of FAIR

The results of FAIR assessments after FAIRification for each of these projects was variable. FAIRplus FAIRification processes yielded very positive results on ReSOLUTE data, achieving a high score in FAIR assessments (83% using RDA indicators v0.03), whereas no improvement in FAIR assessments could be observed in ND4BB data (36% both before and after FAIRplus processes were applied). This can be partly explained by the relative immaturity of FAIRplus FAIRification processes, and partly explained by the choices made by squad teams when prioritising the effort deployed on each dataset. These choices were based on the potential value that could be extracted from each of the four pilot projects, and the effort and complexity involved in applying FAIRplus processes to the data.

CMMI Evaluation

FAIRplus CMMI criteria are being developed for FAIRplus. As they are not yet complete, there is no formal assessment of maturity for the current FAIR cookbook recipes. It is expected that tasks 2.1.4 and 2.2.2 will use the recipes described as part of this deliverable as driving use-cases to help further refine CMMI assessments.

5. Conclusion

Deliverable 3.1 has successfully completed an 11 step “end to end” FAIRification (see Figure 1) for the pilot datasets selected as part of deliverable 1.1, following the current consensus FAIRplus FAIRification process. All four pilot projects have entries in the IMI data catalog, along with an indication of their level of FAIR.

Recipes describing the steps involved in FAIRification have been produced for each project, and will be adapted and utilised against the second phase of IMI projects selected for FAIRification. It is anticipated that the next phase of data will highlight the need to make many adaptations to the current recipes for more general deployment, and we will use the emerging CMMI framework to guide these adaptations and improvements.

The level of FAIR for each of the first four pilot datasets was variable. This highlights that FAIRification processes are still relatively immature. We expect to learn more about the underlying reasons for FAIR improvements over the next 12 months, and we will explore the variability of FAIRification outcomes in more detail in D3.3 (“Report on IMI projects for data types and current technical solutions”, M24). The variability of our outcomes presents a clear goal; to refine recipes (WP2), tools and process implementations (WP3) to increase CMMI levels of FAIRplus FAIRification processes throughout the rest of the project.

During FAIRification of the four pilot datasets, it was identified that the scope for improvement of each project in accordance with FAIR principles was potentially very broad. A key learning from the first phase FAIRification is to identify very clearly defined use cases through active engagement between academic and EFPIA partners that drive development of FAIRplus FAIRification processes. Such use cases will inform squad teams and task teams in work packages 2 and 3, highlighting where their effort can be most effectively invested into the FAIRification process.