IMI2 Project 802750 - FAIRplus

FAIRification of IMI and EFPIA data

# WP2 – Standards definition and process development

# D2.4 FAIR Cookbook - Public Version


| | |
|---|---|
| Lead contributor | Philippe Rocca-Serra (3 – University of Oxford) |
| | philippe.rocca-serra@oerc.ox.ac.uk |
| Other contributors | Susanna-Assunta Sansone (3 – University of Oxford) |
| | Andrea Splendiani (20 – Novartis) |
| | Robert Giessmann (21 – Bayer) |


| | |
|---|---|
| Due date | 30 June 2020 |
| Delivery date | 30 June 2020 |
| Deliverable type | R |
| Dissemination level | PU |

| Description of Work | Version | Date |
|---|---|---|
| | V1.0 | 30 June 2020 |


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.

# Document History

| Version | Date | Description |
|---|---|---|
| V0.1 | 17.06 2020 | First Draft by Dr Rocca-Serra and Prof Sansone |
| V0.2 | 18.06 2020 | Comments and Edits by Robert Giessmann |
| V0.3 | 22.06.2020 | Comments and Edits by Dr Splendiani |
| V0.4 | 30.06.2020 | Comments and Edits by Jan-Willem Boiten, Fuqi Xu, Tony Burdett, Melanie Courtot |
| V1.0 | 30.06.2020 | Final Version |

## Table of Contents

# 1. Executive Summary

The FAIR Cookbook aims to collate protocols for making data FAIR: Findable, Accessible, Interoperable, Reusable. This includes how to FAIRify a number of exemplar datasets, putting the FAIR principles in practice; basic knowledge of what makes data more or less FAIR; understand levels and indicators of FAIRness, and the corresponding maturity model; how to apply the technologies and tools available to assess and improve FAIRness; learn about user roles and their different perspectives on data, and on the challenges when making data FAIR enough for a specific purpose.

Continued efforts by consortium partners, organised in “squads” have significantly improved the breadth and depth of the FAIR cookbook coverage. Between the launch of the FAIRplus cookbook and the second and latest release, 17 new “recipes” have been added.

# 2. Methods

#### Content creation process

The tangible aim is the production of fully documented procedures, solving the use cases assigned to the subgroup. The themes covered by the recipes are defined using both a *top-down approach* and *bottom-up approach*.

The former resulted in the creation of a *prospective table of content,* which identified themes that appeared key to the cookbook developers.

In the latter, content was prioritized based on the needs collected in collaboration with WP1 and EFPIA partners. Briefly, use cases were collected from EFPIA partners or from IMI datasets and triaged for prioritization. Following an *Agile* methodology, organized in 2 squads of roughly equal size and with members spanning diverse skill-sets, FAIRplus cookbook developers, representing a close collaboration between WP2 and WP3, worked under a 3 month-long “scrum” schedule, using weekly calls to assess progress and exchange feedback.

The triage worked by mapping the use cases along the “FAIRification path” outlined in FAIRplus Deliverable 3.1.

These procedures were broken down into the smallest possible units (“recipes”) and revised/reviewed by peers. Slack communication, email and github infrastructure were exploited to the full to meet goals and deadlines.

Various knowledge elicitation techniques were used to capture domain expert knowledge. In that respect, mermaid library has been selected in order to generate standardized graphical representations of the overall key steps of a given recipe.

#### Editorial Process

An editorial board has been established to review the contributions and ensure quality control criteria are met before final integration to the production FAIRplus cookbook.

Review protocols have been developed and structured into a review report form, which encompasses criteria ranging from coverage, syntax compliance, language, code presence and reproducibility/execution. The acronym “calcult” was coined to describe these keys facets of the review process.

The WP2/FAIRplus editorial board also worked on assessing failsafe procedures to limit the effect of technology dependency (jupyterbook and github) by testing and documenting migration of different framework and hosting platforms (Mkdocs and netlify, respectively).

#### Release Process

Continuous integration and release workflows have been set up to automate the tasks and further increase end-user-friendliness. Consequently, the chosen technological solution can be easily replicated by interested persons, is relatively low cost software-wise, and thus presents a sustainable and viable option technically. Keeping its content up to date is, on the other hand, out of scope of this deliverable.

# 3. Results

In an initial phase, FAIRplus cookbook developers have decided to invest in the platform and process, to facilitate scalability and maintainability, by evaluating different platforms (jupyterbook, mkdocs…) and hosting services (github, netlify). This has resulted in the creation of deployment workflows, executed each time a contribution is sent to the book’s content, and ensuring the cookbook output is always in sync with the latest text additions.

In terms of direct visible output, 17 new recipes have been developed, 5 of which contain executable code in the form of jupyter notebooks or workflows.

During this period of development, a particular focus has been placed on discoverability and interoperability, thus with an emphasis on data catalogue and ontologies.

The FAIRPlus Cookbook is currently hosted on github with a production and development servers:

prod: https://fairplus.github.io/the-fair-cookbook/intro

dev: https://fairplus.github.io/cookbook-dev/intro

The source code is available from:

https://github.com/FAIRplus/the-fair-cookbook

Official releases will be made and digital object identifiers will be issued via the github site and its zenodo integration.

Furthermore, based on initial user tests and users feedback, the recipe layout and rendering have been modified to accommodate this user experience, thereby greatly improving readability and accessibility.

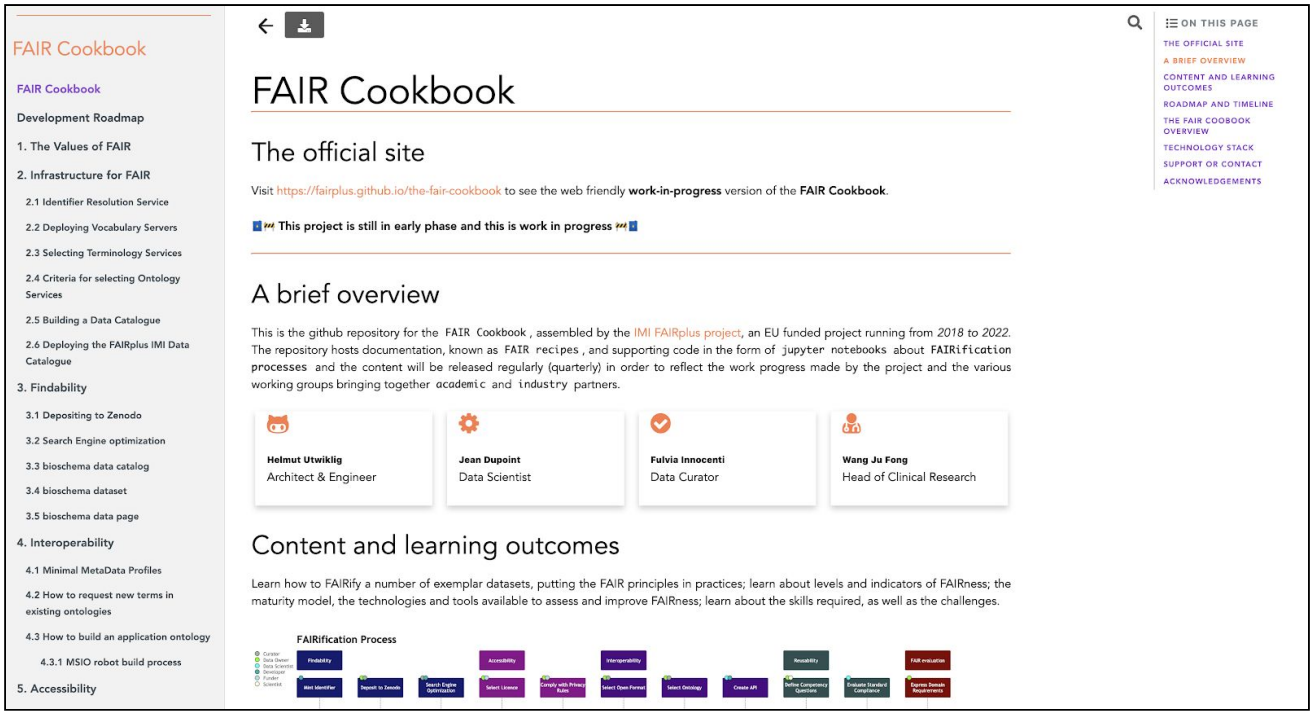
Initial feedback from users highlighted that recipe layout and rendering were not optimal; we have consequently focused towards improving readability and accessibility.

A dedicated CSS styling layer has been applied on top of the “plain vanilla” jupyterbook output. This greatly improved the appeal and appearance of this FAIRplus product, thereby providing a much stronger web presence to the project.

In the next iteration, we will explore opportunities to have a more complete UX review of the cookbook to improve users’ journey.

Work is underway to integrate the elements of Capability Maturity Model Indicator (CMMI) , another product of the FAIRplus project, as a means to document how to foster change in organizations.

Finally, the FAIRcookbook overall layout and table of contents is being used to define the curriculum of the FAIRplus fellowship program in collaboration with the FAIRplus WP4.



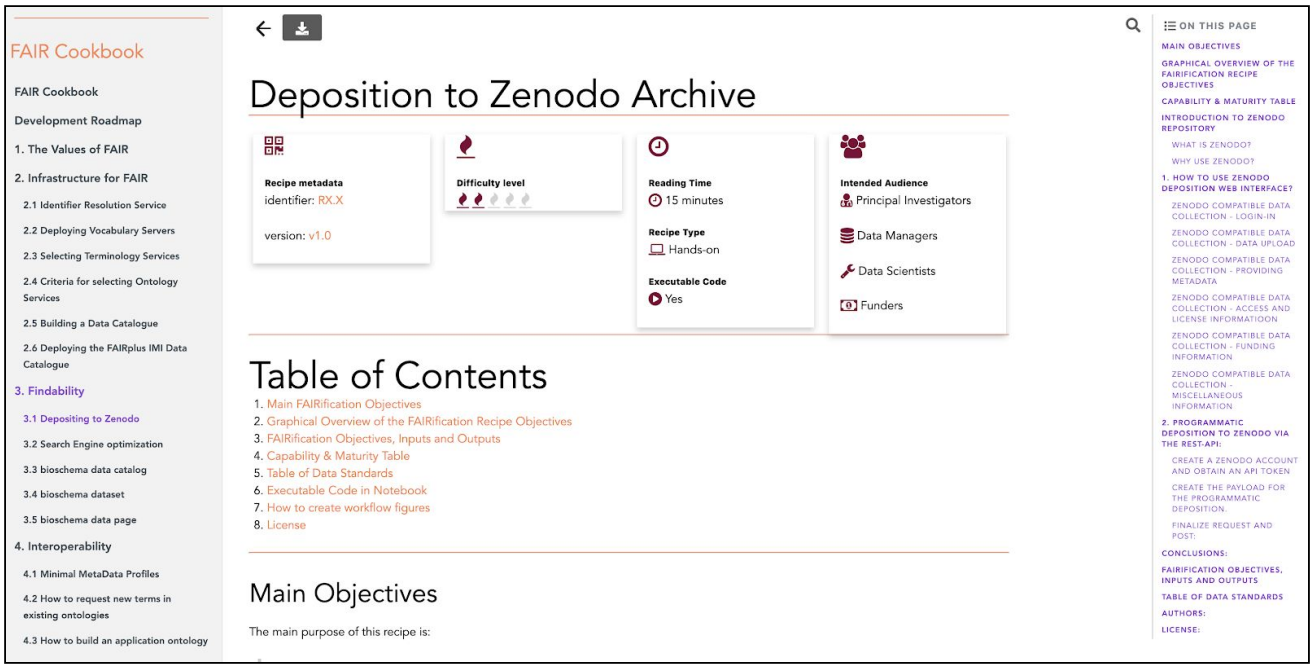



Fig1. Upper section: the welcome page of the FAIRcookbook showing various *personas.* Lower section: a representative recipe, covering an action to address Findability, showing the standardized summary & navigation 'cards' and overall guidance.

# 4. Discussion

Progress on the cookbook has been solid and steady, and productivity is increasing as authors, editors and developers become more familiar with the technologies, but also as the FAIRplus project itself yields more outcomes. The overall FAIRplus process can be streamlined: from use case description to recipe writing, review and publication, decreasing the turnaround time

Productivity tools such as HackMD for redaction of markdown documents, as well as mermaid flowchart editors for knowledge elicitation have demonstrated value in speeding up content creation.

The "cookbook community" is currently looking into further possibilities to improve FAIRnessof the cookbook itself. This includes categorization in respect to different skill levels, user personas, or specific problems at hand.

# 5. Conclusion

The first 6 months since the initial release have seen huge progress, both in terms of content growth and operational streamlining. The FAIR cookbook has the potential to establish itself as an authority and key resource to the IMI organization and the projects it supports for diverse approaches to make data more "FAIR". Beyond the IMI,

outreach and collaboration with projects such as NIH Common Fund Data Exchange and the Pistoia Alliance, which develops the “FAIR toolkit” is underway and will hopefully bring further benefits.

Our work has also identified critical points which represent bottlenecks or potential risks to the cookbook:

- **Need for a strong web presence.** To increase breadth of outreach of the cookbook and maximise its usefulness, we need to ensure it can benefit from a refined web presence. This would help ensure it figures prominently in results from search engines for example, and reaches a wide range of potential users. This would benefit from improvement to the user experience we describe above, and we will investigate the opportunity to receive feedback from web designers.
- **Need for active community contribution.** We have worked extensively towards producing workflow simplifying contribution paths to the cookbook. Encouraging participation remains a high goal or the cookbook team.