

IMI2 Project 802750 - FAIRplus
FAIRification of IMI and EFPIA data

WP2 – Standards definition and process development

D2.1 FAIR Cookbook

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) Wei Gu (7 – University of Luxembourg) Danielle Welter (7 – University of Luxembourg) Tooba Abbassi Daloi (5 – Maastricht University) Laura Portell-Silva (8 – Barcelona Supercomputing Centre)

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

Description of Work	Version	Date
	V0.1	30 June 2022

Document History

Version	Date	Description
V0.1	14 June 2022	First Draft by Philippe Rocca-Serra and Susanna-Assunta Sansone
V0.2	17 June 2022	Comments and edits by Wei Gu, Danielle Welter, Tooba Abbassi Daloi, Laura Portell-Silva
V0.3	24 June 2022	Reviews Tony Burdett, Alasdair Gray
V1.0	30 June 2022	Final Version

Table of Contents

Document History	2
Executive Summary	3
Introduction	4
Methods	5
Technical infrastructure - overview	5
Framework and syntax	5
Visual identity	6
Citability of recipes and credit to authors	6
Integration with other key FAIRplus outcomes	6
Search engine optimisation	7
Operations	7
The FAIRness of the Cookbook	8
Content creation process	8
Editorial process	8
Release process	9
Results	9
Use and adoption	9
Role in ELIXIR	10
Discussion	11
Conclusion	12
Repository for primary data	12

1. Executive Summary

The FAIR Cookbook is an online resource with a focus on the Life Sciences with content (a collection of “recipes”) that help users to implement FAIR principles and/or improve the FAIRness of their data. It is a resource for ‘FAIR doers’ and its objectives are to:

- guide *researchers* and *data stewards* in their FAIRification journey, providing hands-on and technical step-by-step examples to follow;
- provide *policy makers* and *trainers* with practical examples to recommend in their guidance and use in their educational material;
- contribute to enabling the necessary culture change to deliver FAIR in organisations;
- filling the current gaps between high-level FAIR principles and actual implementation.

The FAIR Cookbook currently serves over 70 citable recipes, which provide a broad and in-depth coverage of the technical operational processes allowing to deliver FAIR datasets and build FAIR capacity, as well a set of exemplar FAIRified datasets. The credited contributors include almost 100 researchers and data professionals across pharma and academia, including ELIXIR members not directly involved in FAIRplus. The FAIR Cookbook is offered as a JupyterBook¹, deployed and hosted on GitHub, with an ELIXIR sub-domain name². This infrastructure provides operational processes from community building to technical procedures such as deployment, identifier minting and editorial decision

The FAIR Cookbook integrates the dataset maturity model developed as part of Deliverable D2.2 as well as the Pistoia Alliance’s FAIRtoolkit. Work is in progress to cross-reference recipes to other complementary resources in the ELIXIR ecosystems, such as FAIRsharing (for standards and databases), RDMkit (for high-level signposting of research data management resources).

The FAIR Cookbook is an ELIXIR service provided by the UK and the Luxembourg Nodes. Other Nodes are working on adding this resource to their Service Delivery Plan, a step that will contribute towards the sustainability of the FAIR Cookbook as part of the ELIXIR FAIR Service Framework, part of the ELIXIR Interoperability Platform. The membership of the Editorial Board also reflects the role that ELIXIR Nodes play for the future of this resource.

¹ <https://jupyterbook.org/en/stable/intro.html>

² <https://faircookbook.elixir-europe.org/>

2. Introduction

Created by researchers and data managers professionals, the FAIR Cookbook is an online resource for the Life Sciences with recipes that help you to make and keep data FAIR. The FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. However, the FAIR Principles are aspirational and generic. The FAIR Cookbook guides researchers and data stewards of the Life Science domain in their FAIRification journey; and also provides policymakers and trainers with practical examples to recommend in their guidance and use in their educational material. The FAIR Cookbook provides recipes to learn: how to FAIRify datasets, the levels and indicators of FAIRness, the maturity model, the technologies, the tools and the standards available, as well as the skills required, and the challenges, to achieve and improve FAIRness (Fig 1).

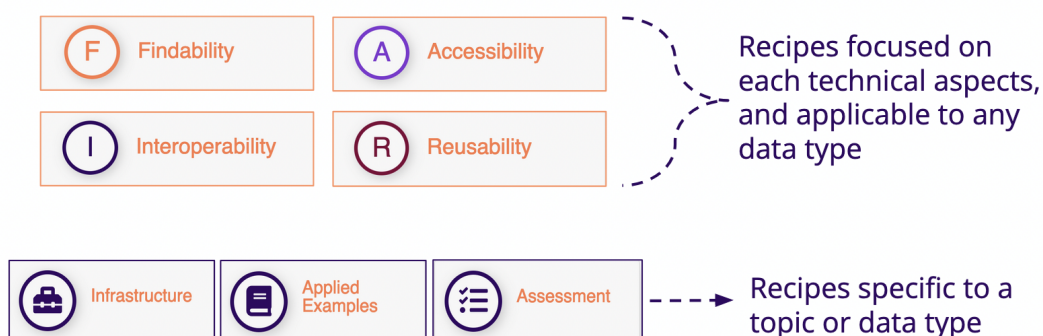


Figure 1. Type of recipes showing the types and what elements they cover and address.

The FAIR Cookbook organises the recipes according to the FAIR elements, audience type (your role), reading time, and level of difficulty (Fig 2). The FAIR Cookbook is a live resource, meaning that content is added and improved, iteratively, in an open manner, therefore bear with us if several sections are still work in progress.



Figure 2. Example of some of the recipes that assist users in meeting each of FAIR elements.

3. Methods

Technical infrastructure - overview

The FAIR Cookbook technical infrastructure is built on open source community practices, using *jupyter book*, following the practice used by The Alan Turing Institute's "The Turing Way Book of Data Science"³, an open source community-driven guide to reproducible, ethical, inclusive and collaborative data science.

The technology stack includes: GitHub for version control and hosting; jupyterbook engine; written materials in markdown; HackMD markdown editor, integrated with GitHub; jupyter notebooks for executable code; binder for the web execution of jupyter notebook distributed with a recipe; mermaid javascript library for flowcharts, Gantt charts and pie charts.

Framework and syntax

For the content management, we use jupyterbook which allows content to be written in Markdown. To ease engagement with contributors, the HackMD⁴ extension for the Chrome web browser was used to write, edit, and collaboratively review content generated by the contributors. Content styling was aligned with the Alan Turing Institute's book of data science style guide⁵. Computational notebooks in the form of native jupyter notebooks are seamlessly integrated in the jupyterbook infrastructure and can be executed on either Binder⁶ or Google Colabs⁷ cloud infrastructure at the click of a button.

³ <https://www.turing.ac.uk/research/research-projects/turing-way>

⁴ <https://hackmd.io>

⁵ <https://the-turing-way.netlify.app/community-handbook/style/style-crossref.html>

⁶ <https://mybinder.org/>

⁷ <https://colab.research.google.com/>

Visual identity

The custom cascading style sheets (CSS) used to render Myst Markdown panels as cards and icons are provided by fontawesome 5.0 (free and open version). The colour scheme is aligned with that defined by the FAIRplus consortium. Illustrations are generated using the mermaid javascript library to produce user journeys, Gantt charts, class diagrams and graphs. The mermaid live editor⁸ was used to draw collaboratively during knowledge elicitation steps. Images were standardised and files saved as mmd files before being rendered to png.

Citability of recipes and credit to authors

Following the evaluation of various persistent identifier services, (e.g. handle.net, doi, and w3id), the w3id⁹ approach was selected based on the absence of costs, relative ease of use and cost effectiveness. A process is being worked out to automate w3id minting via a GitHub action. In this way the persistent findability of the recipes is enabled. Furthermore, to incentivize contributions and credit contributors for each recipe we (recommend the) use of ORCID and then we attribute each author via CreDiT ontology (Figure 2).

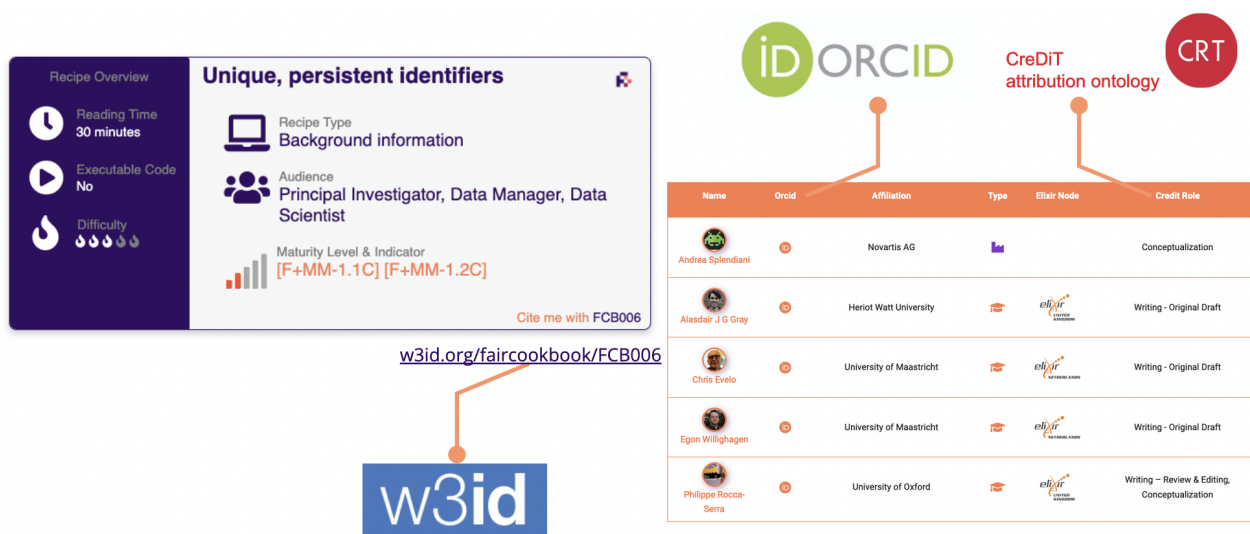


Figure 3. Citability of recipes and identification of and credit for authors.

Integration with other key FAIRplus outcomes

The recipes now also integrate another key output from a WP2 FAIRplus work stream: the Data Set Maturity (DSM) model and its dedicated GitHub repository¹⁰. As part of the editorial review process, the DSM indicators related to the recipe are included to

⁸ <https://mermaid-js.github.io/mermaid-live-editor/#>

⁹ <https://w3id.org>

¹⁰ <https://fairplus.github.io/Data-Maturity/>

specify which maturity level could be achieved for that particular aspect by following the full extent of the recipe. Depending on the recipe type, one or more DSM indicators may apply (Figure 4). Reviewers need to be trained or familiarised with the DSM to ensure consistent assessment. The information about DSM levels and indicators is added to each recipe card and work is ongoing to organise navigation of the book based on these indicators.

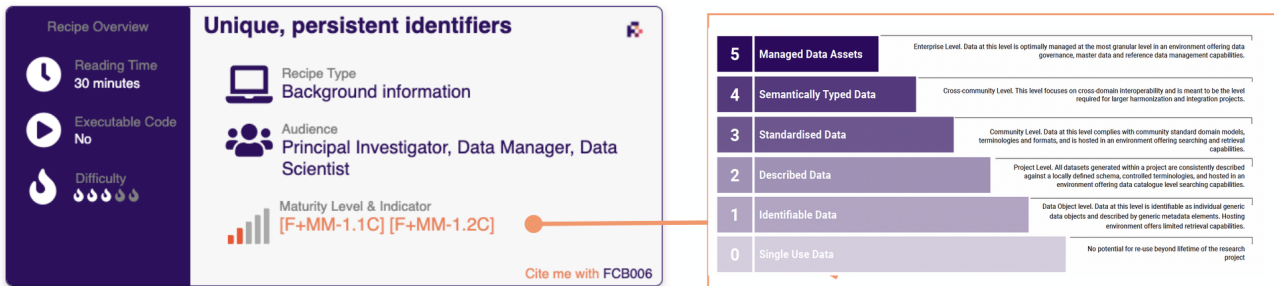


Figure 4. Recipe card showing the Data Set Maturity Indicators, providing information about the maximum level of maturity that may be reached if applying the recipe. In this instance, the indicators 1.1C and 1.2C cover aspects of data “Content”, in contrast to Hosting (H) and Representation (R).

Search engine optimisation

The jupyterbook infrastructure provides minimal support for allowing content to be indexed by search engines. The default jupyterbook was augmented to rely on the sphinx-sitemap extension to generate the sitemap.xml file requirement by search engines to crawl a site. In addition, each recipe content page is marked up via custom python extension code. The code harvests key metadata found in the recipe’s backbone structure to populate a general machine-readable description of the recipe using the Bioschemas Training Material profile¹¹ and the detailed steps using the *Schema.org* HowTo Type¹². This results in a JSON-LD payload being added to the HTML head section of each recipe page. These profiles are validated using Google Structured Data Tool¹³.

Operations

Two groups have been devised to supervise the development of the FAIR Cookbook. A Steering Group providing scientific oversight and an Editorial Board focussing on day to day operations. Fortnightly one-hour calls are held to track progress and raise issues while monthly content creation events are scheduled and aligned with GitHub milestones to ensure at a pace development.

¹¹ <https://bioschemas.org/profiles/TrainingMaterial/1.0-RELEASE>

¹² <https://schema.org/HowTo>

The FAIRness of the Cookbook

As an online resource, the FAIR Cookbook is also FAIR by the use of the following technologies and standards:

- Findability:
 - in search indexes via sitemap.xml and JSON-LD; markup of the recipes with schema.org and Bioschemas conforming to a community agreed standard;
 - citable recipes via w3id.org unique persistent identifiers;
 - credit to authors via ORCID (where available).
- Accessibility:
 - via HTTPS protocol.
- Interoperability:
 - JSON-LD markup;
 - cross-link to objects in other registries;
 - attribution to authors via CreDiT ontology.
- Reusability:
 - Creative Commons Attribution 4.0 International (CC BY 4.0) licence.

Content creation process

The tangible aim is the production of fully documented procedures. The themes covered by the recipes are defined using both a *top-down* 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 prioritised 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 prioritisation. Following an *Agile* methodology, organised 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 gathering techniques were used to capture domain expert knowledge. In that respect, the mermaid library has been selected in order to generate standardised 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. The reviews encompass criteria ranging from coverage, syntax compliance, language, code

presence and reproducibility/execution. The Editorial Board also worked on assessing failsafe procedures to limit the effects of technology dependency (jupyterbook and GitHub) by testing and documenting migration of different frameworks 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 easily be 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.

4. Results

Use and adoption

Launched in early 2020, the FAIR Cookbook has quickly attracted new contributors and users, quite successfully considering the specialised nature of its content. The latest statistics shows that in the past 24 months, 22,000 users accessed the FAIR Cookbook (Figure 5).

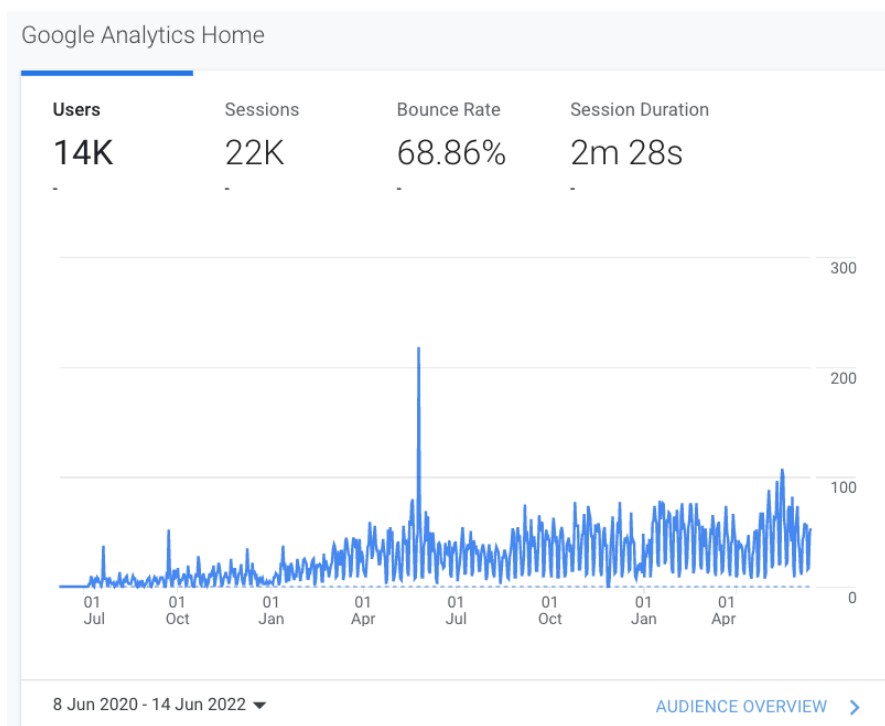


Figure 5. Google analytics trend since launch. The audience spike corresponds to the Elixir Webinar in May 2021 and one can see the consistent traffic growth trend as more content was added to the FAIR cookbook.

In May 2021, an ELIXIR-organised webinar on the FAIR Cookbook¹⁴ attracted almost 300 attendees (out of over 400 registrations), of which more than half were academics, the rest from the private sector; this is the highest number and webinar attendance rate recorded by ELIXIR to date. A recent invited presentation to the USA NIH Office of Data Science Strategy's FAIR WG (formed by NIH science officers and programme managers) has led to an open discussion on how the NIH can contribute and participate in the growth and evolution of the FAIR Cookbook.

At European Commission (EC) level, the FAIR Cookbook is recommended in the IHI (previously IMI) Project Guidelines for Open Access to Publication and Research Data¹⁵, in the "IMI Tools Catalogue"¹⁶, and it features in the EU Innovation Radar Platform¹⁷. The FAIR Cookbook is also a recommended resource in the EC report on "Maximising investments in health research: FAIR data for a coordinated COVID-19 response"¹⁸, and in the EOSC FAIRsFAIR's "FAIR Adoption Handbook & Report on Good Practices in FAIR Competence Education"¹⁹.

A recently signed MoU between ELIXIR FAIRplus and the Pistoia Alliance focuses on the FAIR Cookbook, reinforcing the value and interest of the pharma R&D and training groups for this resource. The MoU also outlines a path to further coordinate content development and cross-references between the FAIR Cookbook and the FAIR Toolkit²⁰.

Role in ELIXIR

The FAIR Cookbook also features as one of the achievements in the ELIXIR Annual Report for 2020, as a new component of the Interoperability Platform²¹. Now it is also a core named element the ELIXIR Interoperability Platform 2022-2023 programme's two tasks, namely:

- The ELIXIR FAIR Service Architecture Framework, which will serve to demonstrate impact, benefits and actions via "interoperability stories" that are recipes in the Cookbook; and
- The ELIXIR Knowledge Hub, which will be based on the FAIR Cookbook and RDMkit²², as well as on ELIXIR registries, such as FAIRsharing, to collate and disseminate the outputs and knowledge, also generated by the aforementioned

¹⁴ <https://elixir-europe.org/events/fairplus-webinar-discovering-fair-cookbook>

¹⁵

https://www.imi.europa.eu/sites/default/files/uploads/documents/resources-for-projects/IMI2_OpenAccessGuidelines_Updated2021.pdf

¹⁶ <https://www.imi.europa.eu/projects-results/catalogue-project-tools>

¹⁷ <https://www.innoradar.eu/innovation/39384>

¹⁸ <https://data.europa.eu/doi/10.2777/726950>

¹⁹ <https://doi.org/10.5281/zenodo.5787046>

²⁰ <https://fairtoolkit.pistoiaalliance.org>

²¹ <https://elixir-europe.org/sites/default/files/documents/annual-report-2020.pdf>

²² <https://rdmkit.elixir-europe.org/>

framework and across the other ELIXIR platforms.

This participation in the ELIXIR Interoperability Platform is also facilitate the work towards progressively cross-referencing recipes to other relevant ELIXIR registries and other resources in the FAIR ecosystem, such as FAIRsharing (for standards), bio.tools (for tools), Workflow Hub (for workflows), as well as the RDMkit (for the research data management cycle).

The FAIR Cookbook is an ELIXIR service provided by the UK and the Luxembourg Nodes. Other Nodes are working to add this resource to their Service Delivery Plan, a step that will contribute towards the sustainability of the FAIR Cookbook. The FAIR Cookbook has been designed from the start for long term sustainability by the ELIXIR Nodes, with contributions via a private-public-partnership from key collaborators, and some support from the ELIXIR Hub, via embedding it into future ELIXIR programmes. The current achievements are evidence that the FAIR Cookbook has already impacted and shaped ELIXIR activities, and this will have ramifications into its communities of practice. Having become a resource supported by several Nodes, the FAIR Cookbook now also received a new ELIXIR sub-domain name²³.

5. Discussion

Sustainability is always a primary concern, but we are tackling this from four fronts: infrastructure, content, embedding and endorsements. The technical infrastructure is lightweight and off the shelf with a light hosting footprint. However, the UI and searchability of the content must improve, and contributions from Bayer and University of Oxford are set to focus on this task for the remainder of the project. Other planned improvements include categorization in respect to different skill levels, user personas, or specific problems at hand, using the EOSC Terms4FAIRskills terminologies²⁴.

The primary sustainability task is content management, which is a distributed responsibility across the contributors and the Boards, whose membership is now primarily with ELIXIR node representatives. As with the majority of the ELIXIR resources, the FAIR Cookbook will be sustained by the Node membership, including those participating in the current ELIXIR platforms, communities and projects, as well as the external partners; responsibility for the FAIR Cookbook lies with the Node governance already in place. Crosslinking recipes to other resources will help to increase traffic and value. Work is also in progress to streamline the content submission process: from use case descriptions to recipe writing, review and publication, decreasing the turnaround time. Productivity tools such as HackMD for creation of markdown documents, and mermaid flowchart editors for knowledge

²³ <https://faircookbook.elixir-europe.org>

²⁴ <https://terms4fairskills.github.io>

elicitation have demonstrated value in speeding up content creation.

The FAIR Cookbook has also succeeded in activating relevant discussions within organisations on the necessary internal steps required to enable capacity building in FAIR data, supported by the essential cultural changes. Recent discussions have highlighted the need for pharmas to create an internal version of the FAIR Cookbook, in sync with the public instance, to add organisation-specific recipes, especially on security and data access topics. The concept of a network of internal versions and the shared common instance of the FAIR Cookbook will be further explored to support content expansion of the public instance, as well as its growth, use and sustainability in a private-public-partnership context.

6. Conclusion

The FAIR Cookbook is a ‘living resource’ with a growing user-base and a thriving community of Life Science professionals who work collaboratively and pre-competitively on describing common exemplar processes and paths to FAIRification, by contributing recipes (adding, reviewing and improving them) iteratively, and in an open manner. Launched in 2020, the FAIR Cookbook has filled a glaring gap in the FAIR-enabling ecosystem of resources: the FAIR Principles are aspirational, and the guidelines available to date were very generic and insufficiently detailed to explain how to deal in practice with the diversity of data types and the complexity of the real scenarios in the Life Sciences.

The FAIR Cookbook has quickly become a catalyst for cooperation on turning FAIR into reality, among researchers and data managers, in the academic and the industrial sectors, across the wider ELIXIR network, and beyond. It has matured quickly enough to have already impacted and be embedded in ELIXIR and other international activities.

7. Repository for primary data

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

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