The FAIR Cookbook –

the essential resource for and by FAIR doers

Philippe Rocca-Serra[1,2]* (0000-0001-9853-5668), Wei Gu[3] (0000-0003-3951-6680), Vassilios Ioannidis[4] (0000-0002-4209-2578), Tooba Abbassi Daloii[5] (0000-0002-4904-3269), Salvador Capella-Gutierrez[6] (0000-0002-0309-604X), Ishwar Chandramouliswaran[7] (0000-0001-9697-9599), Andrea Splendiani[8] (0000-0002-3201-9617), Tony Burdett[9] (0000-0002-2513-5396), Robert T. Giessmann[10,11] (0000-0002-0254-1500), David Henderson[10] (0000-0002-6433-200X), Dominique Batista[1] (0000-0002-2109-489X), Allyson Lister[1] (0000-0002-7702-4495), Ibrahim Emam[12] (0000-0002-7561-2787), Yojana Gadiya[13] (0000-0002-7683-0452), Lucas Giovanni[5] (0000-0001-6518-9535), Egon Willighagen[5] (0000-0001-7542-0286), Chris Evelo[5] (0000-0002-5301-3142), Alasdair J. G. Gray[14] (0000-0002-5711-4872), Philip Gribbon[13] (0000-0001-7655-2459), Nick Juty[15] (0000-0002-2036-8350), Danielle Welter[3] (0000-0003-1058-2668), Karsten Quast[16] (0000-0003-3922-5701), Paul Peeters[17] (0000-0001-9915-2933), Herman van Vlijmen[17] (0000-0002-1915-3141), Tom Plasterer[18] (0000-0002-1527-3381), Dorothy Reilly[19] (0000-0002-6677-3132), Eelke van der Horst[20] (0000-0002-8777-5612), [the FAIR Cookbook Recipes' Authors](), *and Susanna-Assunta Sansone[1,*] (0000-0001-5306-5690).*

* Corresponding authors: [philippe.rocca-serra@oerc.ox.ac.uk](), [susanna-assunta.sansone@oerc.ox.ac.uk]()

Affiliations

1	Oxford e-Research Centre, Department of Engineering Science, University of Oxford, 7 Keble Road, OX13QG, Oxford, UK.
2	AstraZeneca, Data Office, Data Science & AI unit R&D, 136 Hills Rd, Cambridge, UK.
3	Luxembourg Centre for Systems Biomedicine, ELIXIR Luxembourg, University of Luxembourg, L-4367 Belval, Luxembourg.
4	Vital-IT Group, SIB Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland.
5	Department of Bioinformatics (BiGCaT), NUTRIM, FHML, Maastricht University, Maastricht, the Netherlands.
6	Barcelona Supercomputing Center (BSC), Plaça Eusebi Güell, 1-3, 08034-Barcelona, Spain.
7	Office of Data Science Strategy, National Institutes of Health, 9000 Rockville Pike, Bethesda, Maryland 20892. USA.
8	IQVIA, Kirchgartenstrasse 14, Basel, Switzerland.
9	European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, CB10 1SD, UK.
10	Bayer AG, Business Development & Licensing & OI, Pharmaceuticals, 13342 Berlin, Germany.
11	Institute for Globally Distributed Open Research and Education (IGDORE), Berlin, Germany.

12	Data Science Institute, Imperial College London, William Penney Laboratory, South Kensington Campus, London, SW7 2AZ, UK
13	Fraunhofer Institute for Translational Medicine and Pharmacology and Fraunhofer Cluster of Excellence for Immune Mediated Diseases, Schnackenburgallee 114, 22525 Hamburg, and Theodor Stern Kai 7, 60590 Frankfurt, Germany.
14	Department of Computer Science, Heriot-Watt University, Edinburgh, EH14 4AS, Scotland, UK.
15	The University of Manchester, Department of Computer Science, The University of Manchester, Manchester, M13 9PL, UK.
16	Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Straße 65, 88397 Biberach an der Riss, Germany.
17	Janssen, Turnhoutseweg 30, B-2340-Beerse, Belgium.
18	AstraZeneca Pharmaceuticals, 36 Gatehouse Drive, Waltham, MA 02451, USA.
19	Novartis Institutes for BioMedical Research, Novartis Pharma AG, Basel, Switzerland.
20	The Hyve BV, Arthur van Schendelstraat 650, 3511 MJ Utrecht, The Netherlands.

Abstract

The notion that data should be Findable, Accessible, Interoperable, Reusable, according to the FAIR Principles, has become a global norm for good data stewardship, a prerequisite for reproducibility. Nowadays, FAIR guides data policy actions and professional practices in the public and private sectors. However, despite such global endorsements, the FAIR Principles are aspirational, remaining elusive at best, and intimidating at worst. To address the lack of practical guidance, and help with capability gaps, we developed the FAIR Cookbook, an online resource of hands-on recipes for "FAIR doers" in the Life Sciences. Created by researchers and data managers professionals in academia, (bio)pharmaceutical companies and information service industries, the FAIR Cookbook covers the key steps in a FAIRification journey, 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 data FAIRness.

Introduction

The FAIR Principles [1] have taken the world of scientific data management by storm, aligning stakeholders behind a common set of guidelines to ensure data (including datasets and software) is Findable, Accessible, Interoperable, and Reusable. The FAIR Principles describe characteristics for systems and services to make data a valuable asset, which could be rigorously evaluated, extensively reused, reproduced or repurposed, ensuring appropriate credit and legitimate use, to the benefit of both creators and users. The great merit of the FAIR movement has been to propel key, basic concepts of good data management worldwide, gaining acceptance in organizations from all sectors, from academia to industry, from funding agencies to scholarly publishers.

In the Life Sciences, where FAIR was born, public and private organizations strive to implement the principles into practices to unlock data's potential for future research. Major research infrastructure and governmental organizations, such as the pan-European ELIXIR (https://www.elixir-europe.org)

and the USA NIH Office of Data Science Strategy (https://datascience.nih.gov), work to develop and fund resources, practices and policies that enable the collection, storage, analysis, use, and sharing of data according to FAIR practices. Major (bio)pharmaceutical companies value their data as an essential corporate asset, and consider FAIR as a key driver of their digital transformation; companies, providing services and solutions for research, also display FAIR in their portfolio [2,3]. Undoubtedly, FAIR has become essential in support of today's data-driven scientific research, which has shifted towards automation, requiring the use of data at scale by machine, with minimal human intervention. In today's world, large cohort studies and multimodal datasets are routinely produced and used, and the criticality of their data readiness, interoperability, and reusability has been acutely felt during the public health emergency of 2020-2021.

Despite the wide mobilization and adoption, there are two key challenges we must overcome to make FAIR a reality. Firstly, within any organization or project, often it is exceptionally difficult to plot a path to a FAIRer state. The FAIR Principles are aspirational guidelines, and not a formalized standard; also there is neither one specific technology to implement, nor one magic solution to follow. This means that FAIR is not the destination, but it is a continuum, a journey [4], and each path to FAIRness can follow a different route. Also, there is a large body of generic FAIR guidance, which is high level and domain agnostic [5,6,7], but lacking practical examples on "how to" for different data types and scenarios. Coupled with a digital skills shortage and talent gap to be filled, in all sectors, especially around research data stewardship, these factors make it difficult to confidently devise methods that will ensure data will be FAIR. Secondly, it is challenging to accurately evaluate the costs and benefits of FAIR data. To encourage budget holders to invest in FAIR, it is necessary to demonstrate the value of any proposed operational changes associated with the delivery of FAIR data and services; success stories, however, are usually anecdotal. Fostering change in culture around research practices [8,9] depends on a clear statement of the types of improvement that can be achieved as well as on understanding when 'FAIR enough' has been reached, following a cost/benefit analysis [10–12].

The FAIR Cookbook (https://faircookbook.elixir-europe.org) is our contribution towards addressing these two challenges. Launched in 2020, the FAIR Cookbook was created collaboratively and pre-competitively by academics, major (bio)pharmaceutical companies, and information and service companies partners in FAIRplus (https://fairplus-project.eu). This is an international project funded by Innovative Medicine Initiative (IMI, https://www.imi.europa.eu now Innovative Health Initiative, IHI), the largest private-public European partnership program funding health research and innovation. In this paper, we present the FAIR Cookbook, its creation and content, its value, use and adoptions, as well as the participatory process, and collaborative plans for sustainability, which have already moved beyond the initial project.

Results

Use it, adopt it, join it: A practical guide in the FAIRification journey

Created by researchers and data managers professionals from the public and private sectors, the FAIR Cookbook is an online resource that delivers content ("recipes") to help users make and keep data FAIR, with a specific focus in the field of Life Sciences. The technical infrastructure, detailed in the Method section, is built on open-source community practices. The recipes cover the operation steps of FAIR data operations and services. The FAIR Cookbook helps users achieve and improve data FAIRness, by learning how to FAIRify datasets, use levels and indicators of FAIRness, discover the appropriate technologies, tools and standards, as well as understand the skills required, and the

challenges. The FAIR Cookbook is a live resource because recipes are added and improved, iteratively, and peer-reviewed in an open manner.

The two key objectives of the FAIR Cookbook are to fill the current gaps between high-level FAIR Principles and their actual implementation and contribute to the necessary culture change to deliver FAIR in public and private organizations. The FAIR Cookbook is for all stakeholders in the data life cycle in Life Science, who aim to put FAIR into practice, as summarized in Box 1.

- To researchers and data stewards:
    - It provides introductory material and hands-on, technical step-by-step examples for an effective FAIRification journey.
- To policy makers and trainers:
    - It delivers practical examples to recommend in policies and use in educational material to incentivize and guide FAIR in practice.
- To data managers and technical professionals:
    - It offers a venue to document and share existing and new approaches or services to support FAIRification, via recipes that show exemplar FAIRified datasets, FAIRification tools, workflows, and other resources.
    - It also promotes a participatory culture that enables sharing of expertise by getting exposure and credit and contributes towards building organizations' FAIR capacity.

Box 1. The FAIR Cookbook is a resource for everyone.

Live recipes: A combination of guidance, hands-on, and examples

As of August 2022, there are over 70 production-grade recipes in the FAIR Cookbook. The FAIR Cookbook recipes are a combination of guidance, technical, and hands-on, and are organized in two main groups. The first is designed around the FAIR Principles, covering technical and operational processes and applicable to any data types. The second covers three specific topics: (i) software infrastructure to build FAIR processes, such as ontology browsers or data catalogs; (ii) the process of assessing FAIRness using well known tools [13,14]; and (iii) examples of FAIRified datasets in the molecular, pre-clinical and clinical areas. The latter were developed working with data producers in IMI projects and (bio)pharmaceutical companies and provide users with exemplar cost-effective FAIRification pathways in specific contexts.

The concept of a recipe was first prototyped and published as an article [15], and subsequently converted into a proper record in the FAIR Cookbook (http://w3id.org/faircookbook/FCB037). Persistent unique identifiers are used to cite recipes and credit their authors, and the content is released under an open license. Graphical representations are used, where needed, for illustrative purposes. Links or cross-links to complementary online resources provide additional reading material, as relevant. About a fifth of the current recipes have associated executable code, which can be run on cloud-based infrastructures. All technical features are detailed in the Method section.

In each recipe, a summary card displays the reading time, level of difficulty, and audience types, as detailed in the Method section. For instance, recipes for researchers include those on how to choose licenses for datasets, package data for transfer, and provide sufficient descriptive metadata. Recipes for data stewards and managers include those providing exemplar FAIRification datasets, and how to deal with sensitive, patient-centric data. Recipes for technical professionals include those focusing on a number of FAIR-enabling tools and services, and on how to manage terminologies and data curation services. The recipe's summary cards also display maturity levels and indicators, described

in the Method section, which helps users target a maturity level that aligns with an anticipated FAIRification goal. This feature enables the users to assemble recipes, building a coherent path, which follows the data management expectations in terms of FAIRness.

Which recipes: Where and how to start

The FAIR Cookbook is designed to be useful to a variety of users and scenarios. When users seek information on a specific topic, they can browse and select relevant recipes via the search function (https://faircookbook.elixir-europe.org/content/search-wizard.html). The FAIR Cookbook also includes (in the forewords sections) recipes that provide more general guidance, for example, on how to prioritize which project, process or data to FAIRify (https://w3id.org/faircookbook/FCB055), or considerations on what metadata is and how many types there are (https://w3id.org/faircookbook/FCB068). Once users know what their goals are, they can focus on relevant recipes in the sections on the FAIR components, as illustrated in Box 2, or in the sections on the software infrastructure, the assessment tools, and the exemplar FAIRified datasets.

- **Goal**: improving visibility of content
    - Recipes: on identifiers (https://w3id.org/faircookbook/FCB006), also specific to chemical structures (https://w3id.org/faircookbook/FCB007) and search engine optimization (https://w3id.org/faircookbook/FCB010), under Findability.
- **Goal**: semantic integration of datasets from multiple sources
    - Recipes: selecting terminologies and ontologies (https://w3id.org/faircookbook/FCB020) and an ontology service (https://w3id.org/faircookbook/FCB004), under Interoperability.
- **Goal**: security compliance and with regulators
    - Recipes: on secure data transfer protocol (https://w3id.org/faircookbook/FCB014) under Accessibility;
    - declaring data's permitted use (https://w3id.org/faircookbook/FCB035) under Reusability.

Box 2. Examples of goals and relevant recipes that assist towards meeting the needs.

If users need generic guidance on the FAIRification processes, in a sister paper [16], we describe a framework that helps them to: (1) set realistic and practical goals; (2) examine data, capability and resource requirements; (3) assess, design, implement and reiterate; (4) review against the initial goals. This FAIRification process is also outlined in a recipe (https://w3id.org/faircookbook/FCB079).

Contributions and synergies: Cultivating the collective knowledge

As of September 2022, the contributors to the FAIR Cookbook are almost 100 researchers and data managers professionals from more than 30 academic and industrial organizations, across the wider ELIXIR network, and beyond. This ensures diversity and coverage of topics and skills. All authors are FAIR doers, for example, being involved in data management projects, or being in charge of technical or curatorial resources, or being the provider of such services. Our editorial and operational processes, which includes book dash events, and a number of routes to contribute content asynchronously, are detailed in the Method section.

We built a thriving contributing community by leveraging on three motivational points: (i) the need to stay engaged and updated with the latest developments; (ii) the value of being visible in the FAIR space, and recognized as an expert; and (iii) the importance of expanding the network of collaborators, clients, or users. In many cases, the creation of recipes offered opportunities to unlock the collective knowledge towards synergistic approaches. For example, different groups and

projects come together to harmonize the practices around the FAIRification of observational studies and databases, which were captured as a recipe (https://w3id.org/faircookbook/FCB054).

The content creation process was mainly informed by the needs of IMI data-producing projects and the participating (bio)pharmaceutical companies. For example, colleagues at Novartis AG [17] contributed insights to ensure that Contract Research Organizations (CROs) deliver data that is as FAIR as possible to the paying contractees; the result was a new dedicated recipe (http://w3id.org/faircookbook/FCB056). Another specific contribution, however, came from the members of the FAIRplus fellowship Programme (https://fairplus-project.eu/get-involved/fellowship), in the form of improvements to the recipes. This training program for 20 individuals (from academia and industry) was set to empower them to become FAIR ambassadors in their respective organizations. Using their own datasets, and with their own goals and challenges, the FAIRplus fellows reviewed the FAIR Cookbook to identify gaps in coverage, or areas where content needed improvements or extensions. For instance, a deeper coverage of the notion of "provenance information" was requested, along with clarifications about the different types of metadata: structural metadata, administrative metadata, provenance metadata or quality metadata. The provenance and metadata aspects were improved, and covered by two specific recipes (https://w3id.org/faircookbook/FCB036 and https://w3id.org/faircookbook/FCB068), respectively. Current topics of interest and focus are recipes around knowledge graphs, applications of semantic web standards, and more examples of FAIRified datasets.

Value and use: Educational, every-day tasks and strategic purposes

The latest statistics shows that in the past 12 months (August 2021-August 2022), 11,500 users accessed the FAIR Cookbook, totaling 40,200 page views (source Google Analytics). More importantly, we validated the utility and value of the FAIR Cookbook based on three uses: (i) as an educational resource on FAIR in a training context; (ii) as practical guidance on how to improve day-to-day tasks for FAIRer data; and (iii) as a contributor towards changing the culture in research data management. In the sections below, we summarize our findings.

The Cookbook was used as educational material in the context of the FAIRplus fellowship programme; by design fellows were both consumers of existing recipes, and producers of novel ones. The Cookbook helped them gain a deeper understanding in terms of FAIR know-how, and apply the gained expertise to their own data, and in real scenarios. The fellows largely reported that they achieved the expected results, and their experience was a direct observable confirmation of the validity of the recipes' content towards the intended (learning) objectives. There is no denying that some recipes require a greater amount of technical background knowledge, and a steeper learning curve, a fact that is illustrated on recipe cards, at the top of each recipe, as detailed in the Method section. Due to the wide variety of backgrounds among the FAIRplus fellows, some technical content was indeed challenging for some; however, working in pairs helped to optimize information sharing and the learning process. Following this successful experience, we are in the process of embedding the FAIR Cookbook into other relevant programs set to 'training the trainers', such as the FAIR Data Steward Ambassadorship program, part of the ELIXIR-UK Node (https://elixiruknode.org/projects/elixir-uk-dash).

Working with the partners in FAIRplus, we assessed the utility of the FAIR Cookbook in (bio)pharmaceutical companies, where the shift of the data management culture to a FAIR practice represents one of the biggest internal data science bets for these enterprises. The assessment, as well as any subsequent operations, happened behind their firewalls. Therefore, the three outcomes

we report here were expressed in terms of satisfaction of the value of the recipes, against specific tasks, or challenges addressed. In the first example, colleagues at Janssen reported a positive contribution of the FAIR Cookbook towards their discussion on return on investment to operationalize FAIR. Janssen's specific interest was in project asset continuity, and any associated metric that allows to gauge the sustainability and impact of their projects. In particular, implementing procedures that improve, in measurable ways, data asset mobilization and reuse, justifying the necessary efforts and investments. The result was the development of internal tools and a catalog to support data science visualization and reporting, enable cross functional leveraging of Janssen assets internally, and enable reproducibility and interpretability of machine learning models. In the second example, collaborators from Boehringer Ingelheim reported a positive impact of the FAIR Cookbook. They demonstrated internally the power of using open ontologies for performing data integration tasks (https://w3id.org/faircookbook/FCB022), and of building application ontologies (https://w3id.org/faircookbook/FCB023) using an open-source tool and a set of semantic resources, which are open and interoperable by design, from the Open Biomedical Ontologies Foundry [18]. In the third example, colleagues at AstraZeneca requested guidance and best practice to make metadata (data usage rights) FAIR, for an implementation based on DCAT (https://www.w3.org/TR/vocab-dcat-3/); this was achieved via a representation based on the Resource Description Framework standard. The result was the joint creation of a dedicated recipe on digital rights (https://w3id.org/faircookbook/FCB035), which now influences clinical data usage policy across the company internally.

Adoptions and collaborations: A flagship resource in Life Science

We were also successful in reaching a considerable level of international support for and endorsement of the FAIR Cookbook by major research infrastructures, pre-competitive initiatives and governmental organizations, in a very short time. This also reflects the quality, maturity and unique value of this resource. In particular, at European Commission (EC) level, the FAIR Cookbook is recommended by the IMI/IHI Project Guidelines for "Open Access to Publication and [19] Research Data", and by the EC report on FAIR data for a coordinated COVID-19 response [17], among others.

Featured as one of the achievements in ELIXIR Annual Report in 2021, in the same year, the FAIR Cookbook became an ELIXIR service provided by the UK and Luxembourg Nodes. In 2022, the ELIXIR Spain and Switzerland Nodes also started the process of adding the FAIR Cookbook to their service delivery plan. Becoming a Node-provided service is the result of a formal review process that looks at the quality, maturity, use and impact of the proposed resource, and this is also the ELIXIR's formal instrument to ensure the collaborative maintenance and sustainability of highly valuable shared resources. Nowadays, the FAIR Cookbook is among the ELIXIR flagship resources (https://elixir-europe.org/what-we-offer/guidelines).

Working as part of a growing ecosystem of FAIR resources, we also established collaborations with other initiatives and integrated its content in diverse online guidelines. For example, we started to progressively create bi-directional links between sections of the step-by-step instructions in the recipes to other reference material, such the RDMkit (https://rdmkit.elixir-europe.org/index.html), another ELIXIR resource for the data management practices, and the Pistoia Alliance's FAIR Toolkit (https://fairtoolkit.pistoiaalliance.org/), an industry oriented, pre-competitive entry-level guidance to FAIR. Other global resources are referenced via mono-directional links, for example, recipes links to records in the multidisciplinary FAIRsharing [20] (https://fairsharing.org) for more details on standards (reporting requirements, terminology artifacts, models, formats, identifier schemas),

repositories (databases and knowledge bases) and data policies. Work is in progress also to tag recipes with a terminology that describes the competencies, skills and knowledge associated with making and keeping data FAIR (https://www.eoscsecretariat.eu/cocreating-eosc/terms4FAIRskillsPOC).

The FAIR Cookbook was designed from the start for long term sustainability by the ELIXIR network, with contributions via private-public-partnerships, and key collaborators from other major Life Science organizations. This strategy for sustainability is yielding. One notable example is the collaboration with the USA NIH Office of Data Science Strategy, where the Program Director, who oversees the establishment of a FAIR data ecosystem, has become a member of the FAIR Cookbook's Editorial Board. This can potentially enable global knowledge exchange to enhance the impact of the FAIR Cookbook recipes, allow for the exploration of joint activities, such as book dash events, and the development of best practice recipes from USA experts to grow the network of FAIR doers.

Discussion

Created and managed by professionals who work every day with data in academia, (bio)pharmaceutical companies and information service industries, the FAIR Cookbook is a live, open and collaborative resource, which documents the fundamental processes and capabilities for provisioning FAIR data and services. The citable recipes are developed collaboratively by credited authors and anchored to real examples and use cases by data-producing projects and organizations. We illustrated how the FAIR Cookbook is uniquely positioned not only to serve as practical guidance to improve every-day tasks, but also to contribute to a curriculum on FAIR data and inform discussions around the necessary changes to deliver FAIR within organizations. The FAIR Cookbook is a resource for all stakeholders involved in the data life cycle. No single group can cover all topics, because "it takes a village" [21] to deliver FAIR in practice. That is why we have also devised ways to ensure anyone can highlight missing topics, and other experts can contribute, as detailed in the Method section, to the expansion of the content in terms of breadth and depth, as well as filling gaps.

From the onset, we looked at a multi-layered approach to ensure the sustainability of the FAIR Cookbook. As for any resource developed by funded projects, the risk is to be short-lived; long-term sustainability is the common challenge, regardless of how successful a resource becomes. Our approach to sustainability for FAIR Cookbook follows four fronts: (i) infrastructure, (ii) content, (iii) embedding and (iv) endorsements. The technical infrastructure is lightweight, off the shelf with a light hosting footprint, and the code is open; the website has an ELIXIR sub-domain name. The infrastructure's key strengths are its low maintenance and FAIRness as an online resource, as detailed in the Method section. Key challenges are the implementations of new functionalities, but these could be created as part of newly funded projects, which the FAIR Cookbook is included in.

As per content sustainability (the most critical aspect), this is currently a distributed responsibility across the contributors and the Editorial Board, whose current membership is with ELIXIR Nodes and NIH representatives. The FAIR Cookbook has quickly become a catalyst for FAIR doers, successfully unlocking the collective knowledge around each FAIR topic. Key strengths helping content sustainability are its specialized and detailed information, which is also crosslinked to complementarity resources, to reduce duplication and increase traffic; and its citable recipes, credited to authors, motivating contributions. Key challenges are that, although virtually anyone can contribute content, authoring and reviewing recipes, as well as keeping them up-to-date, requires

commitment, hands-on expertise, and time. On the embedding and endorsement fronts, the FAIR Cookbook showed a strong start: it is already recommended by funding agencies, and it is rooted in major infrastructure organizations. Its role in ELIXIR will enable the FAIR Cookbook to grow as part of the ELIXIR framework of FAIR services. Its collaboration with the Pistoia Alliance (https://pistoia-alliance.org) will facilitate the continued collaboration with (bio)pharmaceutical companies and companies that provide services and solutions for research. Furthermore, the participation of the USA NIH Office of Data Science Strategy will be pivotal to expand the operations by involving and serving more academics and private partners for the community's good. The key strengths of the embedding and endorsement fronts are that the FAIR Cookbook is a recognized unique resource of high-quality value, anchored to users' needs; and it has already gained the trust of and the support from key international organizations. The key challenges are to establish a governance and coordination structure that scales up and meets the expectations of a growing community of collaborators.

The FAIR Cookbook has also succeeded in activating two main discussions around how to enable expansion and management of the content in specialized domains, e.g., for plant science, and by the private sector, which can share what is pre-competitive but needs to protect what is proprietary. To meet the interest of the specialized communities, we will explore the creation of Domain Boards, which can be responsible for the coordination of the contributions within their domain community. To serve the needs of the private sector and benefit from their contributions, we will explore the concept of a network of FAIR Cookbook instances, with a public instance and internal versions for organization-specific recipes, for example regarding security and data access topics. Both routes (the Domain Boards and the network of FAIR Cookbook instances) will also contribute towards the sustainability challenges and contribute to the previously mentioned sustainability strategy encompassing infrastructure, content, embedding and endorsements.

We firmly believe that the success of the FAIR Cookbook is due to the timely delivery of specialized content, which fills the glaring gap between high-level FAIR Principles and their actual implementation in the Life Sciences. Our recipe for success is based on crediting expertise and contributions, promoting convergence and collaborations, and acknowledging the diversity of approaches towards the desired level of FAIRness. In a relatively short time, the FAIR Cookbook has attracted international contributors and users outside the project that directly funded its creation. What about you? Use it, contribute to it, and share it!

Method

The FAIR Cookbook relies on well-established off-the-shelf solutions, frameworks and open-source community practices to minimize development and maintenance costs. In the following sections, we detail the approach followed ranging from the infrastructure selection and content creation to publication methods.

Technical infrastructure

The Cookbook is built using Jupyter Book [22], following the example of The Alan Turing Institute's "The Turing Way Book of Data Science" [23], an open source community-driven generic guide to reproducible, ethical, inclusive and collaborative data science. The technology stack includes: GitHub (https://github.com) for version control, hosting, continuous integration, and automation of the build (via custom GitHub Actions); Jupyter Book engine (https://jupyterbook.org); written
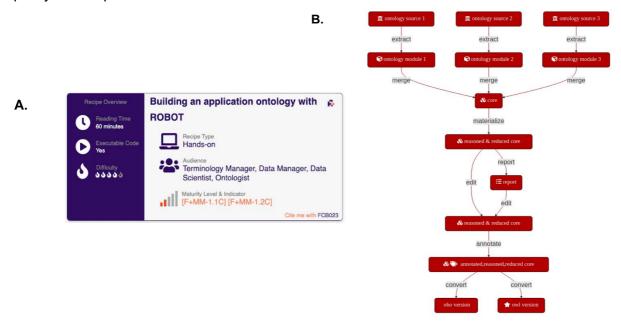
materials in markdown; HackMD markdown editor (http://hackmd.io), integrated with GitHub; Jupyter Notebooks for executable code [24]; binder for the web execution of Jupyter Notebook distributed with a recipe.

Framework and syntax

For content management, we used Jupyter Book which allows content to be written in markdown. To ease engagement with contributors, the HackMD extension for the Chrome web browser was used to write, edit, and collaboratively review the content generated by the contributors. In addition, two more contribution routes were added: a less technically involved path, relying on shared Google documents, and one for technically advanced users who can use integrated development environments to write markdowns or notebooks and to commit to the GitHub repository branches, as they would with code. For both contribution mechanisms, recipe templates were devised. Lastly, computational notebooks in the form of native Jupyter notebooks are seamlessly integrated in the Jupyter Book infrastructure and can be executed on either Binder (https://mybinder.org) or Google Collaboratory (https://colab.research.google.com) [25] cloud infrastructure at the click of a button.

Visual identity and standard layout

Standard web techniques were used to develop a distinctive visual identity for the FAIR Cookbook and sections of the recipes, such as the 'summary cards'; icons were provided by the Font Awesome 5.0 free and open version (https://fontawesome.com). The Mermaid JavaScript library (https://github.com/mermaid-js/mermaid) was used for flowcharts, charts and class diagrams. The Mermaid live editor (https://mermaid-js.github.io/mermaid-live-editor/#) was used to collaboratively create diagrams during knowledge elicitation steps. Images were standardized and files saved as MMD files before being rendered to PNG format. An example is shown in Figure 1. Recipes were structured around a standardized layout, to assist the reader, but also to facilitate the review and quality control process.

Figure 1: (A) A summary card of a FAIR Cookbook recipe (https://w3id.org/faircookbook/FCB023) showing key information. A machine-readable summary metadata is available as a JSON-LD markup embedded in the HTML page. (B) A mermaid generated flowchart diagram providing an overview of the key steps of the procedure described in the recipe, which in this case describes how to build an application ontology.

Citability of recipes and credit to authors

Following the evaluation of various persistent identifier services, (e.g. https://fairsharing.org/search?recordType=identifier_schema), the w3id approach (https://w3id.org/) was selected based on ease of use and cost effectiveness (https://w3id.org/faircookbook/FCB077). Furthermore, to incentivize and credit contributors we enabled the use of ORCID [26], and of the CreDiT ontology [27] to specify author's roles. An example is shown in Figure 2. Furthermore, the FAIR Cookbook resource itself was registered in identifiers.org [28], with its own dedicated namespace (https://registry.identifiers.org/registry/fcb) to enable the content to be replicated in multiple resources under a single unifying URI, facilitating downstream integration.

w3id.org/faircookbook/FCB035

Figure 2. Citability of recipes and identification of and credit for authors; an example is provided.

FAIR Maturity Indicators

The recipes were integrated with the FAIR DataSet Maturity (FAIR-DSM) model [29], another output of the FAIRplus project. The FAIR-DSM indicators specify which maturity level(s) and aspect they measure (content, representation or hosting) may be achieved following the full extent of the recipe; one or more recipes may be needed to fulfill a maturity level. Figure 3 shows how each recipe has been anchored to the FAIR-DSM model and how this is represented. Depending on the recipe type, one or more indicators may apply.
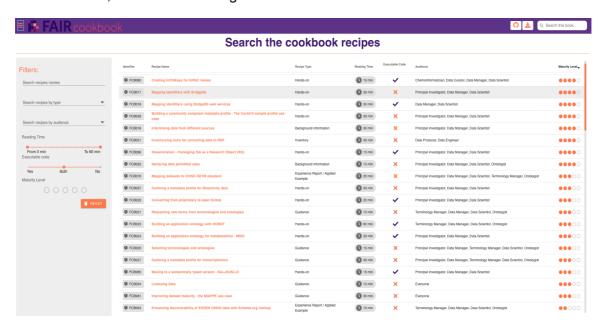

Figure 3. Recipe card showing the DataSet Maturity indicators and corresponding maturity level. In this instance, the indicators 'DSM-4-R6 cover aspects of data "Representation", in contrast to Hosting (H) and Content (C).

The FAIR-DSM model was developed in parallel to the FAIR Cookbook and the integration was realized via a dedicated editorial review process, which required the authors of the recipe to ensure

that the most relevant indicators were used. This integration allows exploration of the FAIR Cookbook also according to the desired improvements.

Search engine optimization and searchability

The Jupyter Book infrastructure provides minimal support for allowing content to be indexed by search engines, therefore, we augmented it relying on the sphinx-sitemap extension to generate the sitemap.xml file, which is required by search engines to crawl a website. Since findability is essential to all online resources, this method is also illustrated in a specific "Search engine optimization" recipe (https://w3id.org/faircookbook/FCB010). In addition, each recipe's content page was marked up via custom Python extension code, which harvests key metadata to generate a machine-readable description using the Bioschemas Training Material profile (https://bioschemas.org/profiles/TrainingMaterial/1.0-RELEASE) [30] and the *Schema.org* HowTo Type (https://schema.org/HowTo). This delivers recipes metadata in JavaScript Object Notation for Linked Data format (JSON-LD, https://json-ld.org) that is added to the HTML head section of each recipe's page. The searchability of the recipes was improved with the inclusion of a dedicated interface and filters, as illustrated in Figure 4.

Figure 4: A searchable view of the FAIR cookbook content available from https://faircookbook.elixir-europe.org/content/search-wizard.html

The FAIRness of the FAIR Cookbook

The FAIR Cookbook is also a FAIR resource itself, and this were made possible by the use of the following technologies and standards:

- Findability: (i) each recipe is identified via a unique persistent identifier provided by w3id.org; (ii) metadata is implemented via standards, such as schema.org and Bioschemas, and the identity of authors is based on ORCID identifiers; (iii) indexing is guaranteed by the search engines optimization via sitemap.xml and JSON-LD.
- Accessibility: via HTTPS protocol.
- Interoperability: (i) JSON-LD markup; (ii) cross-links to records in other registries, as relevant; (iii) attribution roles to authors via the CreDiT ontology.
- Reusability: Creative Commons Attribution 4.0 International (CC BY 4.0) license.

Editorial processes

The development of the FAIR Cookbook is supervised by an Editorial Board, which steers content creation, identifying areas where contributions are needed, tracking progress, and assisting with the review and content production processes. The review process operates very much like that of a scholarly scientific publication. Reviewers are assigned and given a set of criteria to appraise the recipe (including coverage, syntax compliance, language, code presence and reproducibility/execution). Both authors' and reviewers' contributions are credited in the recipe, which receives a unique identifier and is released to production. Members of the Editorial Board also worked with technical staff to ensure failsafe procedures by limiting the effects of technology dependency, testing the migration of different frameworks and hosting platforms. During the initial phase of the work, and to define the content skeleton, we created a Section Board. This was recently rescinded to plan for the creation of Domain Boards, which could coordinate and drive the expert creation of new content specific to Life Science areas.

Content creation and operations in the building phase

To identify the topics for the recipes, a combined *top-down* and *bottom-up approach* was used in the building phase during the FAIRplus project. The former resulted in the creation of a *prospective table of content,* which identified themes that appeared key to the FAIR Cookbook developers. In the latter, the content was triaged for prioritization based on the needs collected from the IMI data-producing projects and the (bio)pharmaceutical companies. The triage worked by mapping the use cases along the FAIRification path [16], and then the procedures were broken down into the smallest possible units to create recipes, which were revised and reviewed. Fortnightly one-hour calls were held to track progress and raise issues, while monthly content creation events, known as 'book dashes', were aligned with GitHub milestones to ensure a steady pace of the development. Major content reviews were performed during quarterly events as part of a group activity done by working under a three month-long "scrum" schedule and using weekly calls to assess progress and exchange feedback. Slack communication, emails and the GitHub infrastructure were exploited to the full to enable and track the discussion. A Code of Conduct is also in place to ensure respectful and healthy interactions. Last, but not least, we created guidelines on what a recipe should or should not be, as illustrated in Box 3.

A recipe should be		A recipe should not be	
Specific	Target a specific task or action or tool and service that brings FAIRness one or more levels up; or provide an example of data FAIRification.	**Too broad and high level**	Should not be a repeat of the full user manual. Should not be a features list of a tool.
Complete	Should be an end-to-end recipe that users can follow and finish a task.	**Incomplete**	Should not be just a teaser that only shows a few steps at the beginning. Should not be an advertisement.
FAIR	The tools and services used should be open, or, if proprietary, a "free" or "community" version should be available.	**Closed**	Users can only test it after purchasing a paid software.

Box 3. Overview principles of what recipes should and should not be.

As we transition out of the FAIRplus project, and enter a maintenance and growth phase, we are working to revise the operations, which will take in account of the Domain Boards, the participation of

the ELIXIR Nodes, as well as the collaboration with the USA NIH Office of Data Science Strategy, the Pistoia Alliance and any other interested organizations.

Release process and containerization

Continuous integration and release workflows were set up to automate the tasks and further increase end-user-friendliness. Furthermore, the integration of Zenodo with the GitHub infrastructure made it possible to automatically generate digital object identifiers with each new version release of the Cookbook. Using the GitHub functionality to cite a code repository independently from each new release, we added metadata formatted according to citation file format (https://github.com/citation-file-format/citation-file-format), which is in compliance with software citation principles [31]. Both aspects complement the stable w3id identifier minting for individual recipes. Finally, cited "collections" of related content can be populated under Zenodo "community" collections for highlighting specific project outputs in accordance with targeted audience findability and reusability. A Docker-based (https://docker.com) version of the FAIR Cookbook is available from the repository, enabling straightforward deployment. The containerization simplifies the interaction and local testing by authors prior to contributing content back to the code repository. It also eases on-premises deployment following content customization as is often the case with industrial partners.

Data availability statement

N/a

Code availability

The FAIR Cookbook is published at https://faircookbook.elixir-europe.org; the content and code are available from a dedicated public repository on GitHub: https://github.com/FAIRplus/the-fair-cookbook, under the CC-BY-4.0 license.

References

1. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
2. Wise, J. *et al.* Implementation and relevance of FAIR data principles in biopharmaceutical R&D. *Drug Discov. Today* **24**, 933–938 (2019).
3. 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).
4. Directorate-General for Research and Innovation (European Commission). Turning FAIR into reality - Publications Office of the EU. *Turning FAIR into reality* http://dx.doi.org/10.2777/54599 (2018).
5. Engelhardt, C. *et al.* D7.4 How to be FAIR with your data. A teaching and training handbook for higher education institutions. *Zenodo* (2022) doi:10.5281/zenodo.6674301.
6. Sustkova, H. P. *et al.* FAIR Convergence Matrix: Optimizing the Reuse of Existing FAIR-Related Resources. *Data Intelligence* **2**, 158–170 (2020).
7. [1902.11162] The FAIR Funder pilot programme to make it easy for funders to require and for grantees to produce FAIR Data. https://arxiv.org/abs/1902.11162.

8. Martone, M. & Nakamura, R. Changing the culture on data management and sharing: getting ready for the new NIH data sharing policy. *Harvard Data Science Review* (2022) doi:10.1162/99608f92.6650ce2b.

9. Bjaalie, J. G., Goble, C., Sansone, S.-A., Nakamura, R. & Martone, M. Perspectives on Data Sharing and the New NIH policy from the European Union. *Harvard Data Science Review* (2022) doi:10.1162/99608f92.bcd0b999.

10. 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).

11. Alharbi, E., Skeva, R., Juty, N., Jay, C. & Goble, C. Exploring the current practices, costs and benefits of FAIR Implementation in pharmaceutical Research and Development: A Qualitative Interview Study. *Data Intelligence* 1–14 (2021) doi:10.1162/dint_a_00109.

12. Alharbi, E., Juty, N., Jay, C. & Goble, C. The FAIR-Decide development. *Zenodo* (2022) doi:10.5281/zenodo.6623516.

13. Wilkinson, M. D. *et al.* Evaluating FAIR Maturity Through a Scalable, Automated, Community-Governed Framework. *BioRxiv* (2019) doi:10.1101/649202.

14. Clark, D. J. B. *et al.* FAIRshake: toolkit to evaluate the findability, accessibility, interoperability, and reusability of research digital resources. *BioRxiv* (2019) doi:10.1101/657676.

15. Rocca-Serra, P. & Sansone, S.-A. Experiment design driven FAIRification of omics data matrices, an exemplar. *Sci. Data* **6**, 271 (2019).

16. Welter, D. *et al.* FAIR in action - a flexible framework to guide FAIRification. *Zenodo* (2022) doi:10.5281/zenodo.7157066

17. Maximising investments in health research - Publications Office of the EU. https://data.europa.eu/doi/10.2777/726950.

18. Smith, B. *et al.* The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* **25**, 1251–1255 (2007).

19. IMI2 project guidelines for open access topublications and research data. *https://www.imi.europa.eu* https://www.imi.europa.eu/sites/default/files/uploads/documents/resources-for-projects/IMI2_OpenAcesGuidelines_Updated2021.pdf.

20. Sansone, S.-A. *et al.* FAIRsharing as a community approach to standards, repositories and policies. *Nat. Biotechnol.* **37**, 358–367 (2019).

21. Borgman, C. L. & Bourne, P. E. Why it takes a village to manage and share data. *Harvard Data Science Review* (2022) doi:10.1162/99608f92.42eec111.

22. Community, E. B. Jupyter Book. *Zenodo* (2020) doi:10.5281/zenodo.4539666.

23. Community, T. T. W. *et al.* The Turing Way: A Handbook for Reproducible Data Science. *Zenodo* (2019) doi:10.5281/zenodo.3381446.

24. Perkel, J. M. Why Jupyter is data scientists' computational notebook of choice. *Nature* **563**, 145–146 (2018).

25. Bisong, E. Google Colaboratory. in *Building machine learning and deep learning models on google cloud platform: A comprehensive guide for beginners* 59–64 (Apress, 2019). doi:10.1007/978-1-4842-4470-8_7.

26. Meadows, A. & Koester, V. ORCID – unique author identifier. *ChemViews* (2015) doi:10.1002/chemv.201500088.

27. McNutt, M. K. *et al.* Transparency in authors' contributions and responsibilities to promote integrity in scientific publication. *Proc Natl Acad Sci USA* **115**, 2557–2560 (2018).

28. Juty, N., Le Novère, N. & Laibe, C. Identifiers.org and MIRIAM Registry: community resources to provide persistent identification. *Nucleic Acids Res.* **40**, D580-6 (2012).

29. Overview | FAIRplus Data Maturity. https://fairplus.github.io/Data-Maturity/.

30. Garcia Castro, L. J. *et al.* Data validation and schema interoperability. (2020) doi:10.37044/osf.io/8qdse.

31. Smith, A. M., Katz, D. S., Niemeyer, K. E. & FORCE11 Software Citation Working Group. Software citation principles. *PeerJ Computer Science* **2**, e86 (2016).

Acknowledgments

Author contributions

- S.A.S. and P.R.S. wrote the manuscript, with input from all authors.
- P.R.S., W.G., V.I., T.A.D., S.C.G., I.C. and S.A.S. are members of the Editorial Board.
- R.T.G., T.B. and A.S. are former members of the Editorial Board.
- P.R.S., D.B., R.T.G., D.H., A.L., I.E., E.W., and S.A.S. contributed to the infrastructure development.
- Y.G., L.G., A.J.G.G., P.G., N.J., D.W. are members of the former Section Board.
- K.G., P.P., H.v.V., T.P. contributed examples of use in pharmas.
- The authors listed in "FAIR Cookbook Authors" contributed recipes.

Competing interests

This manuscript has been submitted to *Scientific Data*. In this context, S.A.S. is Honorary Academic Editor of *Scientific Data* and P.R.S. is a member of the *Scientific Data* Senior Editorial Board.