IMI2 Project 802750 – FAIRplus
FAIRification of IMI and EFPIA data
WP5 – Project management, coordination, dissemination and sustainability

D5.5 Sustainability Plan

Lead contributor	Serena Scollen (01 – EMBL)
	serena.scollen@elixir-europe.org
Other contributors	Nick Lynch (10 – Open PHACTS)
	Bryn Williams-Jones (10 – Open PHACTS)
	Hannah Hurst (01 – EMBL)
	Paul Peeters (16 – Janssen)
	Susanna Sansone (03 – UOXF)
	Philippe Roca-Serra (03 – UOXF)
	Alejandra Delfin (15 – UPS)
	Anne Cambon-Thomsen (15 – UPS)
	Ibrahim Emam (11 – ICL)
	Nick Juty (04 – UNIMAN)
	Andreas Pippow (02 – Fraunhofer)
	Oya Beyan (02 – Fraunhofer)

Due date	31 Dec 2020
Delivery date	18 Dec 2020
Deliverable type	R
Dissemination level	PU

Description of Work	Version	Date
	V1.0	18 Dec 2020

Document History

Version	Date	Description
V0.1	04 Nov 2020	First Draft
V0.2	11 Dec 2020	Document sent for review by the Managing Board
V0.2	15 Dec 2020	Update following the review
V1.0	18 Dec 2020	Final Version

Table of Contents

1. Executive Summary

FAIR data management in itself needs a long term strategy and sustainability plan to be viable. Funders, researchers and organisations all need to incorporate changes to the amount of time and resource spent on managing data to maximise the potential of the data. FAIRplus has committed to open access, open data and open source where applicable to support sustainability.

The FAIRplus project outputs all need different approaches for sustainability. The drive to sustain outputs from this project is largely motivated by the need for continuing improvement and development in this area as well as the need to engage as a scientific community (both as consumers but equally as contributors). This plan for sustainability outlines the options and reasoning for the approaches taken for the FAIRplus project outputs, which present a varied selection of the deliverables. Emphasis has been on the FAIR Cookbook (WP2/3), the Capability Maturity Model (CMM) (WP2/3), the FAIR Fellowship Programme (WP4) and framework on ethics, data and legal implications of data (WP6) as these were viewed by the consortium as the highest priority. This emphasis on outputs with the maximum value for the wider community and potential reuse, has simplified the options for sustainability and allowed focus on routes to uptake with external groups and potential partners. Consideration has been given to how the FAIRplus outputs can be sustained across the academic, biopharmaceutical and SME sectors. There are growing numbers of organisations providing 'FAIR services' either as commercial offerings (SMEs, but larger consulting organisations too) or as part of other FAIR & Research Data Management initiatives. Our long term aim is to allow the FAIRplus outputs to be part of those efforts, and thereby ensure that the users support further enrichment and uptake as they engage in their own FAIR data management.

At this moment, the FAIRplus project has had funding from the IMI and matched in-kind funding from EFPIA partners totalling 8.3M Euros. There is currently no knowledge of funding or future funding (IMI or other EC funded project proposal or a future call for proposals/topics) that would align or be a full extension of FAIRplus or that has a focus on making industry data FAIR. It is therefore deemed unlikely that funding at the scale of this project will be available to continue the activities. Despite this, activities within FAIRplus are contributing to advocacy efforts to make the case for additional funds e.g. general communication/outreach tasks and a FAIRplus webinar 'what is the value of FAIR data' policy event that will be held in December 2020. Using real life examples and use cases, this online event will showcase the value of FAIR data in pharmaceutical research and make the case for policy changes in relation to research data management and research funding. It is aimed at policy makers, data producers and data managers in industry and academia. Additional direct meetings with decision-makers in funding organisations will be organised, providing an opportunity to discuss topics outlined in the FAIRplus Policy Maker Engagement Strategy (D4.4). FAIRplus has also actively interacted

and collaborated with the wider FAIR landscape of projects and initiatives in order to aid alignment, avoid duplication and consider how to move this active community forward.

This sustainability plan is also supported by the Data Management Plan D5.6[1].

2. Methods

The objective of this deliverable was to ensure that all FAIRplus outputs that have been valued as 'assets' are available and used beyond the end of the FAIRplus project.

The following steps have been followed to ensure sustainability of FAIRplus 'assets':

1. Sustain asset identification:
    a. Input was requested from all WPLs at a Management Board meeting and all project participants were contacted by email to contribute
    b. An inventory was generated to collate all possible assets
2. Assess the value of assets to sustain qualitatively:
    a. A sustain value questionnaire was formulated to gather information on the value of each asset
    b. The Work Package or task leads or deliverable owners for each asset completed the questionnaire
    c. Follow up discussions were held when necessary
    d. WPLs discussed the value at a Management Board meeting
3. Steps (1) and (2) were discussed at the FAIRplus Mid-Term Review to consider feedback from the IMI Project Officer, the reviewers and the SIAB and EAB members
4. Develop a sustain plan/pathway for all assets that are considered valuable
5. Open discussions with potential partners and communities to establish and formalise a pathway to sustain assets.

In Scope:
- Assets in scope include FAIRplus methodologies, processes and training for data management.

Out of Scope:
- Preservation of data and sustainability of repositories are considered out of scope for this task. FAIRplus is responsible for the metadata submitted to repositories but is not deemed responsible for maintaining the core infrastructure of the data catalogues themselves as the level of funding required for this is beyond the scope of the FAIRplus project and is not timeboxed within the format of a project. Further details on the data sustainability scope are summarised later in this document.

---

[1] D5.6 DMP: https://drive.google.com/open?id=1GVaYFGOycGVy_PjaaxxEmmV11zChgem6

Other supporting activities across WPs:

- The Policy-maker engagement strategy report (D4.4)[2]
- The FAIRplus webinar 'What is the value of FAIR data' will showcase the value of FAIR data in pharmaceutical research and make a case for policy changes in relation to research data management and research funding. It is aimed at policy makers, data producers and data managers in industry and academia.[3]
- Data Management Plan (D5.6)[4]
- The content of this deliverable will feed into the Sustainability White Paper (D3.8) due in December 2021.

3. Results

The FAIRplus consortium has been committed to open access and open source licences to release tools and standards. All assets have been made publicly available and accessible as early as possible. As an example, the Draft FAIR Cookbook was released in December 2019 (see Milestone 2.2: Draft FAIR Cookbook (alpha version)[5]).

3.1. Summary inventory of FAIRplus assets

Table 1. Summary inventory of FAIRplus assets

Asset	WP
FAIR Cookbook	2/3
FAIR Capability Maturity Model (CMM)	2/3
FAIRplus Fellowship Programme	4
Ethics of data & legal framework	6
Dataset prioritisation process	1
Methodologies for establishing squad teams	2/3
FAIRplus website/Key Performance Indicator (KPI) Dashboard	5
Sustainability Plan	5

---

[2] D4.4: https://zenodo.org/record/3733199#.X71GIlP7QUE
[3] "What is the value of FAIR data?" webinar: https://fairplus-project.eu/get-involved/value-FAIR-data
[4] D5.6 DMP: https://drive.google.com/open?id=1GVaYFGOycGVy_PjaaxxEmmV11zChgem6
[5] M2.2: Draft FAIR Cookbook (alpha version):
https://drive.google.com/file/d/1R33m6O8ETKM-0Eor8RuO7hBvtzlktqqE/view?usp=sharing

3.2. Questionnaire

A questionnaire was developed during Q1 2020 to support the sustainability analysis and, given the wide range of outputs from FAIRplus, to apply some common language and methodology to what could and should be sustainable from the project. Equally, within each of the Work Package teams, the idea of sustainability was introduced to encourage the whole team to think about what happens next and their role in that as organisations and individuals. The high level questions featured in the questionnaire are listed in the table below. They seek to explore some key areas of sustainability including core knowledge, platforms that support the outputs, and gauge early feedback from users and contributors.

The questionnaire was used to assess the FAIR Cookbook, FAIR Capability Maturity Model, the FAIR Fellowship Programme and the framework on ethics, data and legal implications of data.

- Who has the core knowledge? Do we need knowledge transfer?
- Definition of the sustain value of the deliverable beyond the project (summary version)
- Are there any dependencies for software/licensing?
- How extensively have the ideas/recipes been tested/validated?
- Feedback from active users of a deliverable
- Does the delivery platform need to change (be updated) to support sustain? Presentation, user experience
- Are they being taken into other projects by current participants?
- What is the current framework/platform that the deliverable is presented in?
  - Does this offer a sustainable route/platform? Via Crowdsourcing/community engagement?
- Possible sustainability partners or strategies beyond the FAIRplus project lifecycle:
  - Uptake by an institute (e.g. EBI[6])
  - Uptake by a research infrastructure (e.g. ELIXIR[7])
  - Uptake by a relevant organisation e.g. Pistoia Alliance[8]
  - Collaboration with the Alan Turing Institute (ATI)'s 'The Turing Way' - A Handbook for reproducible data science[9]
  - Seek further funding?
  - Crowd maintained?
- Any internal/external considerations for EFPIA?

Figure 1. High level questions featured in the Sustainability Questionnaire

---

[6] EBI: https://www.ebi.ac.uk/
[7] ELIXIR: https://elixir-europe.org/
[8] Pistoia Alliance: https://www.pistoiaalliance.org
[9] 'The Turing Way':
https://www.turing.ac.uk/research/research-projects/turing-way-handbook-reproducible-data-science

In the next section, we describe in more detail the sustainability pathway for some of the key outputs from FAIRplus. Since these outputs are still in progress and at different stages of maturity and external engagement, and with only early usage both within the consortium and beyond, the plans listed below are current approaches, but will remain adaptable to more interaction and engagement.

We have created three tiers to express and summarise the current stage of sustainability for each of these outputs (Table 2).

Table 2. Tiered structure to outline stage of maturity within the sustainability process for each output

	Tier		
	Limited solutions found, resourcing unknown, long term partners unknown	**Solutions are in progress, some resourcing options in discussion, long term partner options under consideration**	**Solutions planned and implemented, resourcing in place and long term partners defined**
FAIR Cookbook		✔	
FAIR CMM	✔		
FAIRplus Fellowship Programme		✔	
Framework on ethics, data and legal implications of data		✔	
Dataset prioritisation process			✔
Methodologies for establishing squad teams			✔
FAIRplus website/ KPI Dashboard			✔
Sustainability Plan			✔

3.3. Sustainability pathway for the FAIR Cookbook[10]

The FAIR Cookbook is considered essential to sustain beyond the project duration. By the end of the project it will contain numerous recipes and we predict a user group will be building in the community. It is considered a single point/one-stop-shop for hands-on guidance (for FAIR implementation by reasonably expert practitioners). All content in the FAIR Cookbook is

---

[10] FAIR Cookbook: https://fairplus.github.io/cookbook-dev/intro

licenced under the Creative Commons Attribution 4.0 International (CC BY 4.0)[11] license. The FAIR Cookbook was made open on GitHub as early as feasibly possible at month 12 of the project (as this was not initially anticipated in the Description of Action, a new milestone was added to the project during the Grant Agreement first amendment). This has helped to build the content, feedback and user base. Despite this, it is likely that the user base will need to grow beyond the end of the project to demonstrate the usability and value before it will be funded as a service e.g. within ELIXIR. It also incorporates a technology stack that has been developed.

The future requirements currently envisioned are summarised as follows:
- To be open and collaborative but with clear governance structure
- To have an option to extend (add new recipes, develop further)
- Expertise e.g. platform developers
- Experts to date should be given credit and governance for founding this tool
- Funding to maintain

The dependencies are predominantly around the expert input. Core knowledge to develop the FAIR Cookbook is currently held by project participants e.g. for the technology stack development (University of Oxford and Bayer), community governance and editorial, such as table of content's development and organization (University of Oxford). It needs to be considered that for future development funding opportunities, these experts could be retained or financed to transfer knowledge. There are alternatives for the software and licencing and it is possible to migrate to other platforms if necessary e.g. due to software rot. The FAIR Cookbook currently works on a system where all recipes (except position statements) undergo two rounds of internal reviews and for this rigour to be maintained in the future, resourcing or process is required.

Within the context of the wider FAIR community, and specifically to maintain the flavour of public-private partnership, which is so important to IMI and the participants of the FAIRplus project, we consider a mixed option to be the most appropriate route forward. The reasoning behind this position is outlined in more detail below.

As with so many eScience cornerstones, the FAIR Cookbook needs to continually evolve, yet be in a form which allows participants to both engage with and take the FAIR Cookbook forward into other large projects. Evolving into a wider community effort is key for this to happen.

We feel the most appropriate long term option is for the FAIR Cookbook to become part of the ELIXIR distributed infrastructure for life sciences, possibly as a resource provided by the ELIXIR-UK Node (of which the University of Oxford is a part) in collaboration with other

---

[11] Creative Commons Attribution 4.0 International: https://creativecommons.org/licenses/by/4.0/deed.ast

interested Nodes (e.g. EMBL-EBI). This would enable continued evolution of the FAIR Cookbook as well as participation within wider scientific community efforts. It would allow interested academic parties to further develop the FAIR Cookbook through standard project-based funding mechanisms. One potential benefit could be that within future ELIXIR-driven proposals the scope could be wider than biological data; content could extend and expand to fit the needs of other ELIXIR communities.

It is highly likely, however, that in the near term, the FAIR Cookbook will not have matured enough to be in a position to secure ELIXIR Node/institutional funding, or part-funding, to be taken forward as a service (for example via ELIXIR-UK driven proposals to UKRI, the main UK funding agency for scientific research). Therefore, we see a critical need to maintain the FAIR Cookbook in the shorter term in order for it to mature. We see a partnership between FAIRplus and the Pistoia Alliance[12] as an important route to maintain and further develop the public-private ethos, and bring real world use cases and examples from data consuming industry organisations on an ongoing basis. Whilst in concept, the FAIR Cookbook could be sustained by the Pistoia Alliance, we feel that there are major limitations to this as an approach which would in fact limit sustainability in the longer term:

- Whilst Pistoia Alliance has a high profile in the pharmaceutical industry and associated informatics supply chain, we feel engagement and availability from a wider community is necessary to embed the FAIR Cookbook in the data consuming community as a whole;
- Pistoia Alliance is funded through project-based, specific funding mechanisms. There is a risk that the pharmaceutical industry would not prioritise sustaining/developing IMI project output, which would risk the continued development of the FAIR Cookbook.

To summarise, the preferred route of sustaining this most important FAIRplus project output is a model where ELIXIR (and ELIXIR-UK in particular) supports the FAIR Cookbook as a resource in combination with carefully managed industry partnering mechanisms with the Pistoia Alliance. The only viable route for sustainable maintenance and development of the FAIR Cookbook is through an open and honest broker mechanism such as this.

If it is not possible to maintain and develop the FAIR Cookbook via these routes, then other options may exist to ensure at least some of the content is sustained. One example could be the Alan Turing Institute (ATI)'s 'The Turing Way' - A Handbook for reproducible data science[13], an open source community-driven guide to reproducible, ethical, inclusive and collaborative data science. The goal is a guide for reproducibility, covering version control, testing and continuous integration; expanding in 2020 to cover reproducible research, project design,

---

[12] Pistoia Alliance: https://www.pistoiaalliance.org
[13] 'The Turing Way':
https://www.turing.ac.uk/research/research-projects/turing-way-handbook-reproducible-data-science

communication, collaboration, and ethical research. This is not the preferred option as it is unlikely all the content would be incorporated.

If there is no uptake of the FAIR Cookbook by the end of the project then our University of Oxford partner, also via their participation in the ELIXIR-UK Node, will commit to preserving the pages but would not be able to keep content up to date without financial support.

One factor which is important to understand is the availability of the FAIR Cookbook to SMEs, and the role which they may choose to play in picking up recipes and further developing them alongside their own unique offerings. Again, non-limited wider community ownership and engagement is fundamental to any uptake and usage within the SME community, and again we feel that this is best supported by ownership of the FAIR Cookbook within the wider ELIXIR environment. For SMEs to manage the risk of FAIR Cookbook adoption, it is vital that a clear and sustained future for the FAIR Cookbook is self-evident. There is no business model for an SME owning the FAIR Cookbook in any shape or form – it is far from clear if that were to happen, who would fund it, and how continued public-private engagement would be maintained.

However, there is a business model for SMEs in engaging their clients in FAIRification of data using the FAIR Cookbook to massively simplify and deconvolute their FAIRification activities. This would, of course, be entirely dependent on the ongoing maintenance and development of the FAIR Cookbook through an open and honest broker, as described above.

3.4. Sustainability pathway for the FAIR Capability Maturity Model (CMM)

CMM data FAIRness indicators can serve as a common method for setting FAIR data targets for project funders and data stewards in the life sciences domain and they can be adopted by research data infrastructures and pharmaceutical companies to benchmark their data assets. CMM process areas and maturity levels can be adopted by pharmaceutical companies and organisations which generate biomedical data as a roadmap for change management.

The FAIR CMM will be housed within the FAIR Cookbook to ensure that the material developed within the FAIRplus project is accessible during and beyond the lifetime of the project. Therefore, sustainability relies upon the strategy for the FAIR Cookbook. However, similarly to the FAIR Cookbook, additional funding will be required if there is interest to extend the CMM further. It is intended that the consortium shall publish the work in a peer reviewed journal.

The CMM is not mature enough to finalise a specific sustainability plan at this point, and the associated deliverables are not due until the latter stages of the project. Further discussions

are needed and a decision will be taken between months 36-42. The following options will be considered:

- FAIR Transformation Programme (Fraunhofer)[14]
- Pistoia Alliance[15]
- RDA[16] e.g. RDA FAIR DMM working group[17]
- Incorporation into ELIXIR[18] or other projects e.g. ELIXIR-CONVERGE[19]/others
- National research data infrastructures e.g. NFDI4Health[20] in Germany

Fraunhofer's FAIR Transformation Programme proposal aims to transfer the CMM outputs to an exploitable toolkit and training modules. The programme could be developed and governed by FAIRplus partners. The programme might be composed of modular components specialised in various capabilities, with each module offered by various parties (e.g. SME, academia) who are subject matter experts.

If a formal programme is not developed then a community effort could potentially be established via the other options listed above.

One important factor to keep in mind is that maturity models as a whole in this domain are very much a work in progress and there is a need to demonstrate real world usage through appropriate recipes in the FAIR Cookbook. As such, the development of recipes which exploit the use cases and real operational issues of EFPIA partners in the FAIRplus project, are important. We also need to ensure a continuous engagement with other maturity model efforts in this rapidly moving domain.

3.5. Sustainability pathway for the FAIR Fellowship Programme

By the end of the FAIRplus project, we will have exemplified the value of a FAIR Fellowship Programme, and how important training on FAIRification is to data practitioners.

Training material should be accessible and reusable by FAIR adopters. The online material will be developed with an authoring software called Rise360[21]. The software allows export of learning modules in HTML-format. Rise360 is a commercial software that needs a license for

---

[14] https://docs.google.com/presentation/d/1C6vchZgF88QMfgqmh4045Bh4d3OFCP5EkWNvjItnZ74/edit#slide=id.p1

[15] Pistoia Alliance: https://www.pistoiaalliance.org

[16] RDA: https://www.rd-alliance.org/

[17] RDM FAIR DMM WG: https://www.rd-alliance.org/groups/fair-data-maturity-model-wg

[18] ELIXIR: https://elixir-europe.org/

[19] ELIXIR-CONVERGE: https://elixir-europe.org/about-us/how-funded/eu-projects/converge

[20] NFDI4Health: https://www.nfdi4health.de/

[21] Rise360: https://articulate.com/360/rise

authoring of content. Once exported in scorm or HTML format, it can be viewed on any platform that supports these formats. The goal is to have all online modules published with a CC-BY-SA license. The learning modules will be published on the Fraunhofer learning platform[22] to conduct the training within the project. The system itself will not be available for free beyond the FAIRplus project end. However, the training material component of the Fellowship Programme will be made available on ELIXIR's Training Portal (ELIXIR TeSS[23]), which has a broad user base globally and support via ELIXIR.

To sustain value, it will be critical to explore options to run similar courses after FAIRplus ends to continue to produce certified FAIR data experts. The programme could be converted into a learning system managed by a center for continuing education or an offering such as Coursera[24] or MOOC[25]. A learning management system allows learners to access the online materials in a certain sequence, to ask questions about the material, hand in assigned tasks and get in touch with subject matter experts. Such a system is usually sustained by a centre for continuing education. However, additional funding would be required at a level which has yet to be ascertained. Further to this, these options have a number of perceived disadvantages, such as the additional funding requirements but also the lack of exchange with experts. Therefore, we envision that the contents of the fellowship programme will be incorporated into the Fraunhofer data science training program[26].

We will also consider whether current FAIRplus project partners with institutional interest in the development of value-added course content could sustain the content of the FAIR Fellowship Programme beyond the life of the FAIRplus project or develop an institutional learning programme. It is possible that Fraunhofer will develop this course with a view to targeting pharmaceutical companies, to produce Fraunhofer certified FAIR data experts. It may also be possible to fund via a funder such as the Wellcome Trust. This would be likely to target the audience of the funder, so potentially more of the academic data managers.

It is important to highlight though, that our project deliverables in the FAIR Fellowship Programme are very much viewed as a pilot experiment, and further funding to support this at scale would be needed, and could well form the basis of a future IMI call, should EFPIA decide to prioritise data FAIRification as a future activity. Similarly, the scope and aims of the FAIR Fellowship Programme during the life of the FAIRplus project were seen as experimental, and much further funding may be required to transition this to a sustainable future if that proves to be worthwhile based on our results.

---

[22] Fraunhofer learning platform: https://ilias.fraunhofer.de
[23] https://tess.elixir-europe.org/
[24] Coursera: https://www.coursera.org/
[25] MOOC: https://www.mooc.org/
[26] https://www.bigdata.fraunhofer.de/de/datascientist/zertifizierungen.html

3.6. Sustainability pathway for framework on ethics, data and legal implications of data

The FAIRplus project has to consider Ethical, Legal and Social implications (ELSI) of research data management, especially given that some of the datasets being worked on encompass patient level data (e.g. OncoTrack[27]), as well as provide guidance for future FAIRification activities on how to tackle key ethical and legal issues, creating a framework for advice on data ethics.

The WP6 team's initial focus was on supporting the internal ethical working of FAIRplus on those datasets (clinical data). This has provided valuable insight to what could be future, longer term ethical focused deliverables and sustain options. It will always be mandatory to consider the ethics and legal risks raised by projects to guarantee that data is processed lawfully.

The aim is that the work completed here can be reused. Updates in accordance with the evolving law and ethics frameworks will of course be needed in time.

WP6 has delivered material to support FAIRplus including:

- A procedure to characterise the ethical issues attached to the projects/data sets considered
- A Data Privacy Impact Assessment (DPIA) template
- A set of measures that the persons involved in data stewardship, data governance and data use need to document and implement before FAIRification starts

In order to deal with the potential complexity and breadth of ethical issues, the team has implemented the use of a checklist as the initial way to introduce and breakdown the sections, stages and details of the ethics approach. The checklist covers some of the challenges involved in ethics planning and execution and includes, for example:

- type of data involved (data source and quality)
- kind of persons concerned (patients, minors, healthy people),
- conditions for collection of the data, degree of identifiability etc.
- processes to put in place to ensure respect of individuals'/institutions' rights prior to FAIRification
- required documents to assure compliance (informed consent, ethical approval...)
- responsible persons to involve. They vary between situations and institutions
- legal restrictions regarding the project

---

[27] OncoTrack: http://www.oncotrack.eu/

- organisational and technical (security) measures concerning the data users' access and the use of data

A workflow based on this checklist is to be elaborated. The tool would enable the key sections to be visible to the user based on simple questions (for example, no clinical data, or no genetic data or no minors involved), and hence be directed only to those sections needed to be considered.

The tool is initially aimed for FAIR data practitioners as being a structured way to introduce the ethics deliverables to a wider audience, and to make sure that groups consider data ethics issues at the beginning of a project lifecycle. The checklist allows easy expansion as new items are introduced to the policy, and focuses on usability at its core. We will focus on this usability in the first version over the coming period, including design and communication experience from within and beyond our consortium. The initial investment will be on bringing this checklist first to the consortium for review, and then generating a wider public version.

To maximise the value from this work, effort will be required to make the checklist easy to use or to develop into a platform without losing the necessary precision of the approach. The ethics team has already contributed to the Cookbook highlighting the values underlying or correlated with FAIR principles. This is an important place to introduce ethics early on in the whole FAIR process. Under further consideration for contributing for sustainability and wider dissemination:

- Discuss with ELIXIR/BBMRI to maintain and consider value of alternatives:
- RDA Interest Group on Health – Ethics
- Societal platform (Genotoul ethics and biosciences societal platform, UPS)

There are possible plans to consider other shared ethics deliverables that could include common templates, cross consortia agreements that involve ethics and related issues.

Another aspect aligned to FAIRplus was the required legal agreements needed for working with the data from the IMI projects. We intend to share the learnings from that process as part of the deliverables here to support legal agreement processes in the future.

3.7. Process for prioritisation of data sets and identification of added value created by FAIRification

The dataset selection criteria have been published as a deliverable[28]. As the process will be incorporated into the FAIR Cookbook as a recipe, the sustainability of the prioritisation process

---

[28] https://zenodo.org/record/3596024#.X71I8lP7QUF

is available in the FAIR Cookbook sustainability plan outlined above. The information will also be archived in Zenodo.org. This is a key output to sustain as lessons learnt here can be fed into data management plans of future projects.

3.8. Methodologies for establishing squad teams

Methods were established very early on on how to set up and run the FAIRplus squad teams. This is currently in the format of a README and associated documents/templates. The sustainability plan is to ensure these are hosted within the FAIR Cookbook or on Zenodo.org.

3.9. FAIRplus website and Key Performance Indicator (KPI) Dashboard

ELIXIR will commit to maintaining the website for FAIRplus and static content for at least 1 year after the FAIRplus project ends. The content will then be archived on Zenodo.org.

3.10. Other FAIRplus outputs

Unless confidential, the deliverables, including the Sustainability Plan itself, will be published on Zenodo.org.

4. Discussion

There are several key components to consider:

4.1. Making data FAIR, when applicable, prospectively rather than retrospectively

This requires changes in the data management planning and resourcing at the project planning stage. FAIRplus is holding webinars and has a clear dissemination plan to encourage this.

4.2. Wide community user base and engagement, across a broad range of organisation types/sectors, is key

A broad community user base across the whole public and private biomedical sector is required to drive development of the assets themselves and ensure uptake of results to change the way we manage data. This takes time to establish. To ensure this occurs, the users need to understand what the assets provide. For example, there need to be good entry points and wizards to guide a user through what is available. FAIRplus will be generating these wizards during the second part of the project.

4.3. Mixed model is key to mitigate risk

Some of the suggested sustainability partners are more sustainable themselves than others. For example:

- ELIXIR[29] is the research infrastructure for biological data. ELIXIR is an intergovernmental organisation with 22 member countries plus EMBL. Community driven activities bring together more than 220 research organisations.
- Pistoia Alliance[30] is a global non-profit organisation that is sustained via a funding model with a growing subscription. All but two of the EFPIA partners in FAIRplus (companies with a demonstrated interest in FAIR data management) are members of Pistoia Alliance.

4.4. Components of the project are already being taken into other projects

One example of uptake by other projects is the technology stack for the FAIR Cookbook, which has been reused by the ELIXIR CONVERGE[31] project in the development of its Research Data Management Toolkit.

4.5. FAIRplus is driving forward to engage with key external partners/future partners

Due to an alignment of overarching activities between FAIRplus and Pistoia Alliance, FAIRplus has actively sought engagement with Pistoia Alliance from the start of the project. Representatives from Pistoia Alliance joined our project Kick Off meeting plus the following General Assembly meetings and our SME and Innovation Forum. In addition, ad hoc meetings have been held between FAIRplus and Pistoia Alliance and a joint presentation was given at the FAIR Converge Symposium[32] held in 2020. A formal collaboration (bound by a Memorandum of Understanding) will soon be in place to align activities between the FAIR Cookbook and the Pistoia Alliance FAIR Toolkit[33] activities.

4.6. Considering value proposition for authors and consumers of FAIRplus assets

It will be important to consider the value proposition to the different stakeholders of sustainability – within the FAIRplus outputs there are different roles in the creation and consumption of the outputs. For instance, the present focus for the FAIR Cookbook will be on

---

[29] ELIXIR: https://elixir-europe.org/
[30] Pistoia Alliance: https://www.pistoiaalliance.org
[31] ELIXIR-CONVERGE: https://elixir-europe.org/about-us/how-funded/eu-projects/converge
[32] https://conference.codata.org/FAIRconvergence2020/sessions/185/
[33] Pistoia Alliance FAIR Toolkit: https://fairtoolkit.pistoiaalliance.org/

sustainability via the authors and contributors, since they have the most invested in the work so far. However, it will be critical to enhance the role and engagement of the consumers (and their parent organisations) in the sustainability of the assets. This will be part of our plans going forward, and will be elaborated as we develop the plan fully.

4.7. SMEs will not offer services if there is not significant business potential

Early routes to sustainability should focus and involve the organisational types that are most heavily involved in FAIRplus – the academic collaborators and biopharmaceuticals. For SMEs to be involved in sustainability there needs to be sufficient business potential ('service and product pull') for them to offer services. SMEs are not a reliable route to sustainability as cost/benefit is heavily skewed away from the SME in establishing services, and the attendant risk is high in the absence of clear and committed clients/partners. However support for FAIRplus outputs will come from a wide ecosystem of application providers offering key services and applications at strategic points along the data lifecycle, for instance:

- Ontology management
- Data catalogue
- Experiment capture tools
- Longer term data analysis and data aggregation

4.8. Sustaining data has not been discussed in this plan but is key

As part of FAIRplus, the datasets handled by the project have become more FAIR and therefore more sustainable. Through outreach events, communications channels and the SME and Innovation Forum datasets that have been made more FAIR they are highlighted and given visibility, with the aim to increase the chance of reuse by SMEs, in turn increasing the likelihood of generating further value. However, the aim of the FAIRplus project has always been to demonstrate the potential for increased data sustainability once it has been made FAIR, whereas individual sustainability per data source remains the responsibility of producers and consumers. Whilst it is of course somewhat paradoxical to mention this, it is important that the original scope and intent of FAIRplus is kept in mind – we are focussed on sustaining and developing the methods and process of data FAIRification and not the data themselves, given the huge funding that this could require. These methods of course, can be used by producers as a means of transitioning their data to a more sustainable future.

4.9. Resource Estimates have not been included in these plans

We have omitted estimates of resourcing at this stage since there are several dependencies to be addressed and this depends significantly on which groups are sustaining the resource and

their levels of expertise. We will work these estimates up as we develop the plans and use these estimates to support the decision making process on sustainability options.

5. Conclusion

To manage risk and ensure key outputs are sustained, we have taken a mixed model approach for sustainability. Outputs need to mature, embed and develop further through community use and uptake in future projects and activities. Their future value will be judged by the community who adopts and develops them. Without a user group, and without future funding, the risk remains that the outputs will remain static or will be lost. Even a very low level of maintenance takes resources. For outputs with greater applicability to the scientific community as a whole rather than for a single organisation (e.g. FAIR Cookbook, FAIR CMM), ongoing sustainability and development absolutely requires a flavour of public-private partnership to continue. As such, the mixed model we advocate is a viable route to maintaining that vital wider academic and industry community engagement through inclusion of these outputs in future projects and activities.

During the remainder of the FAIRplus project, we will continue to develop the plans outlined above. In particular, focus will be on securing further interest from EFPIA partners to propose and fund a project to continue activities via the Pistoia Alliance. Similarly, we will explore how ELIXIR, either through the ELIXIR Hub and/or ELIXIR-UK Node, could host and support some of the key outputs in the longer term to ensure continued availability. We will also explore the potential for follow-up future IMI programmes to further develop and exploit activities in the FAIR space. This is, of course, dependent on sufficient interest and commitment from EFPIA companies in the development of such a call topic.

We have identified that several of the WP outputs would benefit from clearer presentation to assist with longer term sustainability. A key element of FAIR adoption and impact is about culture change and so our intention is to focus on these presentational elements that support culture change through provision of these presentation deliverables. We have seen good examples of this from groups in the FAIR ecosystem including the Pistoia Alliance FAIR Toolkit[34]. The group has also been looking at FAIR incentives and learning from the previous approaches to incentives, enforcement measures and their impact. This will be further investigated to support FAIR adoption and sustainability.

Many of the outputs are experimental and/or proof of concepts, so even if successful, routes to individually sustain these outputs would require significant future funding and adoption into other projects and programmes (e.g. the FAIR Fellowship Programme). Whilst the FAIR Fellowship Programme itself necessitates a running project with available experts to be

---

[34] https://fairtoolkit.pistoiaalliance.org/

successful, much of the content and learning from the whole project can be built into other learning programmes in alignment with FAIRplus partners.

Through efforts around the FAIRification of data which may include clinical and/or patient data, we have established a robust checklist-based process which would be reusable in other projects. In these settings it will always be mandatory to consider ethics, and whilst these checklists will of course need to evolve in time as the regulatory and legal environment changes, there are important learnings and approaches that should be sustained alongside the other FAIRplus outputs.

In addition, further discussion with Fraunhofer and potential funding agencies will be necessary to secure an option to develop the training into a more formalised programme. Further to this, we will ensure to raise sustainability at all available opportunities – for example, at upcoming Managing Board meetings, General Assembly meetings and SME and Innovation Forum events. This will ensure that all stakeholders are driving towards the goal of sustainability and ensuring that outputs such as the FAIR Cookbook and FAIR CMM are considered for engagement in future.

SME engagement is an important part of our sustainability plans, although this has to be taken carefully into context. Continued availability and access to FAIRplus output such as the FAIR Cookbook and FAIR CMM will play a key role in FAIR tooling for SMEs in this data space. We should, however, be mindful of the risks involved in any SME adapting to the FAIR space, as this can involve significant cost without any certainty of customers. There are, of course, advantages for SMEs in engaging with FAIR data, but this is best achieved through ongoing projects with their own clients where both parties are engaged with, and committed to, data FAIRification through the use of FAIRplus project outputs. As such, commitment and engagement from EFPIA consumer companies is important, and recognition that SMEs will need help and support to adopt these methods should be kept in mind. We will explore these themes further in the sustainability white paper deliverable (D3.8) due in month 36. Our ongoing SME engagement programmes remain focussed on demonstrating the potential and value of FAIR data to further stimulate discussions with the client organisations.

Ultimately, a commitment to sustaining and changing the way data is managed needs to be driven by a longer term commitment from the European Commission, IMI and EFPIA. As mentioned the possible FAIR incentives and enforcement measures need to be evaluated to reward good data practices. For instance, a radical idea might be to introduce a 'FAIR levy' on future projects where a % of projects' funds are kept back until data meets FAIR indicators. FAIRplus can play a role as an exemplar, but we cannot drive community change from the bottom up without significant further support.

In addition to the Data Management Plan (D5.6) and this Sustainability Plan (D5.5), there will also be a sustainability white paper produced at month 36 (D3.8) that will describe the challenges, learnings and recommendations from the FAIRplus project.