Recommendations for Services in a FAIR data ecosystem

Report by FAIRsFAIR, RDA Europe, OpenAIRE, EOSC-hub, FREYA[1]

Introduction

This report highlights common challenges and priorities, and proposes a set of initial recommendations on how existing data infrastructures can evolve and collaborate to provide services that support the implementation of the FAIR data principles, in particular in the context of building the European Open Science Cloud (EOSC). The report is an output of three workshops[2] designed to explore, discuss and formulate such recommendations and is aimed at stakeholders in the scholarly world and particularly the EOSC Governance.

Gathering recommendations from the community

The workshop series was organised as three half-day events held in April and September 2019, each tailored to different audiences. The workshops examined services in the scholarly and research data ecosystem: what exists, what could be modified, and how can service provisioning could be optimised. These events also provided an opportunity to engage experts and a range of stakeholders on how to turn the vision of FAIR data and services into reality.

The first workshop was targeted at service providers and research infrastructures. At this workshop, three implementation stories were presented on services and initiatives to help make data FAIR, such as the certification of data repositories, services for data management and exploitation, and persistent identifier (PID) services.[3] In break-out groups, workshop participants then discussed challenges and recommendations concerning services to support FAIR data.

The primary audience of the second workshop were research support staff and researchers. Four implementation stories were presented, followed by break-out groups and a panel discussion.[4] The objectives of this workshop were to share perspectives on how to assist researchers with FAIR, to explore existing services and extensions needed to support FAIR research outputs, to understand how services can work together, and to identify further recommendations for supporting FAIR data.

The services presented at these events offered a sample representing the minimum components of the FAIR data technical ecosystem identified in the report on *Turning FAIR into reality*.[5] The presentations and discussions covered the broader scholarly ecosystem, recognising that FAIR data are part of a complex and evolving landscape.

Gaps

During the two workshops, the key needs and areas of improvement were identified by participants. Within the current landscape, some of the biggest gaps include:

1. Lack of a sustainable ecosystem of independent interoperable services with governance, business model(s) and shared responsibilities to support the creation of FAIR research outputs.
2. Need to address equally: 1) the principles related to findability and accessibility which requires mostly technical expertise that can be addressed by generic services (e.g PIDs, cataloguing, discovery and storage); and 2) the principles related to interoperability and reuse which require services that cater to disciplinary needs with specific domain expertise (e.g. ontologies, curation and stewardship provided by domain repositories).
3. Skills and services for data stewardship and preservation are needed to maintain the FAIRness of research outputs over time. Technical and conceptual expertise for data services is necessary.

Recommendations

Suggestions from the first two workshops resulted in an initial set of recommendations for services to support FAIR data[6]. These are collated and stated below, grouped by broad categories:

---

[4] Implementation stories covered Zenodo, FREYA, Wikidata, and CoreTrustSeal. Workshop II report: https://www.openaire.eu/openaire-workshop-making-services-fair-vienna-april-24th-2019
[5] *Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data*, EU 2018. https://doi.org/10.2777/1524
[6] These recommendations were disseminated in the form of a draft report available at https://docs.google.com/document/d/e/2PACX-1vQWf8rX3pPLL1ZTq9RV5Mi6u3X871mj5EnnUM0mXI4orW7gD4KbR8041tOC5L33Oa65650CMUlVtolG/pub

1. Certification:
   a. Certification mechanisms and capability maturity models need to be further developed for and embraced by services to align with FAIR Principles.
   b. Data repositories should undergo FAIR aligned certification such as CoreTrustSeal.[7]
2. Essential infrastructure components:
   Services supporting FAIR data should offer or make use of the following components:
   a. PID services for a wide range of objects, such as publications, researchers, data sets and organisations. Emerging PID types (e.g. for instruments) should be monitored and used when they are mature.
   b. Domain-specific ontologies, as domain-specific requirements have to be taken into account.
   c. Human and machine-readable standards to make datasets findable, reusable and interoperable (licences as one particular example of standards needed for machine readability).
   d. If applicable, metadata that complies with appropriate (domain) standards should be generated and captured automatically (for e.g by instruments).
3. Stewardship:
   To support the effective use and uptake of services enabling FAIR, institutions should:
   a. Establish data stewardship programmes providing simple and intuitive training for researchers, and enable data stewards and researchers who support applications of FAIR.
   b. Support preservation and appraisal of research outputs: Improve and maintain FAIRness of data objects over time and the long-term usability and findability of datasets.
4. Costs:
   a. Determine the cost for services to align with FAIR principles including for data management support, maintenance and long-term preservation.
   b. Develop a sustainable funding model (of services) taking into account that there might be additional costs for FAIR.
   c. Provide support when determining the cost of data management as this is typically underestimated or unknown.
5. Rewards:
   a. Consider level of FAIRness and data sharing as part of research assessment, among other criteria.

---

[7] https://www.coretrustseal.org/

     b.   References to use certified Trustworthy Digital Repositories (TDRs) in Data Management Plans should be recognised and recommended by funders.

6. <u>Collaboration and support</u>:
   a. Set-up and participate in cross-institutional, collaborative communities of practice to advance and implement FAIR services.
   b. Foster global collaboration on FAIR implementation challenges and emerging solutions through organisations such as the Research Data Alliance.[8]
   c. Create practical guidelines on how to enable FAIR in repositories.
   d. Provide skilled legal advisers in institutions to help in preparing robust DMPs.

7. <u>Data management:</u>
   a. There should be a data selection policy that – pre-deposit – recognises that not all research outputs must meet the highest levels of FAIRness, and recognises what has long term value, and has effect immediately after generation.
   b. Data Management Plans should be required early when applying for funding and must have organisational relevance.
   c. Legal aspects should be taken into account from the start of a project.

Prioritisation of recommendations

Following the gathering of recommendations in workshops I and II, the third and final workshop set out to solidify the work and produce outputs to guide the community in the development of services to support FAIR data.[9] The overall approach is illustrated in Figure 1 and may be summarised as follows:

1. Take stock of <u>recommendations</u> gathered so far
2. Assign relative <u>priorities</u> to the recommendations
3. Associate <u>actions</u> to the top-priority recommendations
4. Collect community input on '<u>action owners</u>', i.e. who could take those actions forward
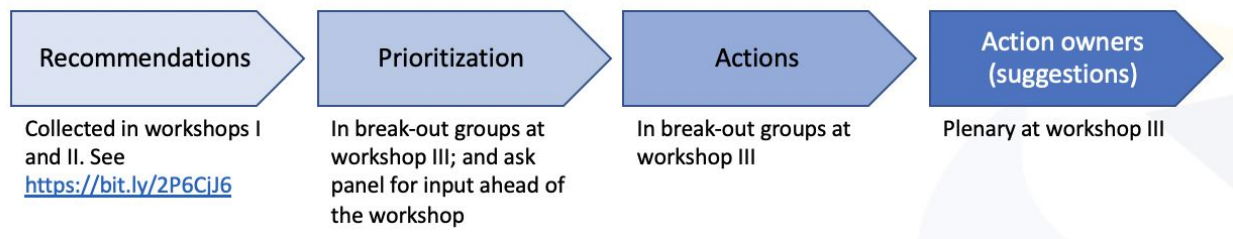

Figure 1: Approach to prioritising recommendations

---

[8] https://www.rd-alliance.org/
[9] Workshop III report: https://www.openaire.eu/blogs/2019-09-30-12-46-02

This section will detail the process that was followed to prioritise the recommendations; actions and action owners will be discussed in the following section.

Prioritisation process

In order to assign relative priorities to each recommendation, we divided the audience into three break-out groups. These groups were chosen to align with different stakeholders: *research institutions*, *service providers* and *libraries*. For each of these, we followed a straightforward ranking exercise: every group member received a total of ten 'votes' which they could freely distribute over the various recommendations. Members were free to give all their votes to a single recommendation, divide their votes over ten recommendations, or anything in between. We then tallied the votes per recommendations to yield a prioritisation score for every recommendation (simply put, most votes meant highest priority score).

For the five panellists, input was collected prior to the workshop in the form of a prioritised ordering of all recommendations. Scores were then assigned according to the priority order (highest score to the top-priority and so on). Averaging the scores across panellists resulted in a priority score for each recommendation from the panel as a whole.

As the panel consisted of five people and the break-out groups consisted of 10-15 each, we recognise that there will be sizable statistical fluctuations in the priority scores calculated this way. To reduce statistical noise, we aggregated data by clustering the recommendations into quartiles, meaning that we only distinguish between four categories:

- First quartile: low priority, denoted by one star (*)
- Second quartile: medium priority, denoted by two stars (**)
- Third quartile: high priority, denoted by three stars (***)
- Fourth quartile: top priority, denoted by four stars (****)

Any conclusions and recommendations in this report are based on these broader categories rather than on the exact priority scores.

Prioritised recommendations

The outcome of the prioritisation exercise is summarised in Figure 2. The various recommendations are displayed as rows, while the different stakeholder groups, plus the expert panel, are distributed over the columns. The color coding indicates the relative priority, from one

star (light red) for lowest priority to four stars (dark green) for the highest. As explained above, the relative priority corresponds to the quartile of the overall vote distribution.

Finally, the Harvey balls on the left of Figure 2 indicate the overall ranking assigned to the recommendation when weighted equally over the different stakeholder groups plus panel (i.e. a full Harvey ball means highest overall priority).

	Recommendation	Research institutions	Service providers	Libraries	Panel
	Certification				
◕	Certification mechanisms and capability maturity models need to be further developed for and embraced by services to align with FAIR Principles	*	**	*	***
◑	Data repositories should undergo FAIR aligned certification such as CoreTrustSeal	***	**	***	**
	Essential infrastructure components				
●	PID services for a wide range of objects, such as publications, researchers, data sets and organisations. Emerging PID types (e.g. for instruments) should be monitored and used when they	**	***	****	****
◕	Domain-specific ontologies, as domain-specific requirements have to be taken into account	***	****	*	****
◑	Human and machine-readable standards to make datasets findable, reusable and interoperable (licences as one particular example of standards needed for machine readability)	*	***	***	**
●	If applicable, metadata that complies with appropriate (domain) standards should be generated and captured automatically (for e.g by instruments)	****	****	**	****
	Stewardship				
◑	Establish data stewardship programmes providing simple and intuitive training for researchers, and enable data stewards and researchers who support applications of FAIR	****	**	**	**
◑	Support preservation and appraisal of research outputs: Improve and maintain FAIRness of data objects over time and the long-term usability and findability of datasets	**	*	****	***
	Costs				
◑	Determine the cost for services to align with FAIR principles including for data management support, maintenance and long-term preservation	**	*	***	***
◕	Develop a sustainable funding model (of services) taking into account that there might be additional costs for FAIR	**	***	***	***
○	Provide support when determining the cost of data management as this is typically underestimated or unknown	*	*	**	**
	Rewards				
●	Consider FAIR compliance and data sharing as part of research assessment, among other criteria	****	**	****	***
○	References to use certified Trustworthy Digital Repositories (TDRs) in Data Management Plans should be recognised and recommended by funders	**	*	*	**
	Collaboration and support				
◔	Set-up and participate in cross-institutional, collaborative communities of practice to advance and implement FAIR services	*	*	**	****
●	Foster global collaboration on FAIR implementation challenges and emerging solutions through organisations such as the Research Data Alliance	****	***	**	****
◔	Create practical guidelines on how to enable FAIR in repositories	*	****	*	*
○	Provide skilled legal advisers in institutions to help in preparing robust DMPs	*	*	**	*
	Data management				
◑	There should be a data selection policy that – pre-deposit – recognises that not all research outputs must meet the highest levels of FAIRness, and recognizes· what has long term value, and	*	****	****	*
◔	Data Management Plans should be required early when applying for funding and must have organisational relevance	****	**	*	*
◔	Legal aspects should be taken into account from the start of a project	*	**	***	*

Figure 2: Prioritised recommendations

As may readily be observed in Figure 2, there is substantial variability between priorities as assigned by the different stakeholder groups. For example, *practical guidelines on how to enable*

FAIR in repositories was seen as a top priority by service providers but as a low priority by the other stakeholder groups and the panel. Similarly, *establishing data stewardship programmes* was seen as a top priority by research institutions but only as a medium priority by the others.

Notwithstanding this variability, four recommendations stand out as being assigned at least medium priority by all, and top priority by two different groups. These are the following:

- PID services for a wide range of objects, such as publications, researchers, datasets and organisations. Emerging PID types (e.g. for instruments) should be monitored and used when they are mature.
- If applicable, metadata that complies with appropriate (domain) standards should be generated and captured automatically (for e.g by instruments).
- Consider FAIR alignment and data sharing as part of research assessment, among other criteria.
- Foster global collaboration on FAIR implementation challenges and emerging solutions through organisations such as the Research Data Alliance.

From recommendations to actions

With recommendations now prioritised, participants in the break-out groups brainstormed about possible actions to implement their top recommendation, considering feasibility. From the actions suggested, again a top action was selected by the different stakeholder groups. It needs to be noted that these actions reflect the discussions in the different stakeholder groups at the time and are not necessarily suitable for generalisation.

The selected priorities and subsequent actions discussed by the breakout groups formed the basis for a discussion on which stakeholders could take on the responsibilities in the services ecosystem for FAIR data for the various actions. To gather input from the audience on possible action owners for the identified actions, we used the interactive presentation tool Mentimeter. See Figure 3 for examples of feedback on two of the actions, one formulated by the libraries group and one by the service providers group.
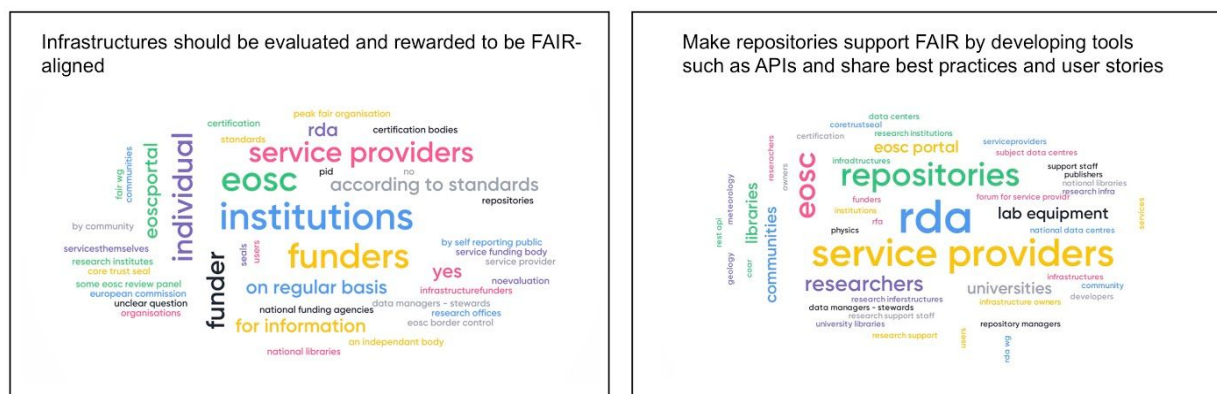
Figure 3: Audience input in the form of word clouds on possible action owners for two of the actions defined by the break-out groups

Table 1 presents the selected priorities by the three stakeholder groups, the matching actions they selected to be most appropriate, and the three most commonly suggested action owners for each action.

Table 1: Priorities, actions and suggested action owners according to stakeholder group

Group	Priority	Action	Suggested Action Owners
Libraries	Consider FAIR alignment and data sharing as part of research assessment, among other criteria.	Infrastructures should be evaluated and rewarded to be FAIR-aligned. Reward researchers who apply the FAIR principles to their research, e.g. through incentives such as increased visibility for their work.	EOSC Funders Service providers Funders Community Universities
Service providers	Domain-specific ontologies, as domain-specific requirements have to be taken into account.	Identify disciplines which don't have ontologies and create awareness for registries of ontologies[10] and enrich them. Make repositories support FAIR by developing tools such as APIs and share best practices and user stories.	RDA Service Providers Repositories

---

[10] The BARTOC registry was specifically mentioned during the breakout groups.

Research institutions	Establish data stewardship programmes providing simple and intuitive training for researchers, and enable data stewards and researchers who support applications of FAIR.	Identify and present the cost of developing supporting infrastructure, including human resources.	Service providers Institutions EOSC

The workshop closed with an open discussion involving the expert panel and audience. A number of additional considerations were raised, such as: the need for EOSC to include an overarching Authentication and Authorisation Infrastructure (AAI); a suggestion to implement highly automatable Digital Object management holistically along the whole research data life cycle; and an encouragement to involve national libraries in the discussions and events around FAIR.

Conclusion

This report presents the outcome of an active process of community consultation - most notably in the form of three workshops held in 2019 - to gather, discuss and analyse recommendations for data services and research infrastructures to support the implementation of the FAIR principles. Coming from a broad range of participants, representing several stakeholder groups, these recommendations provide valuable insights into what the participants perceive to be the greatest impediments, challenges, and opportunities for services to support FAIR data. These insights give further direction and impetus to the development of a FAIR data ecosystem as envisioned in the *Turning FAIR into reality* report, in particular in the context of building the European Open Science Cloud. To deliver tangible and actionable results, with a view of facilitating adoption, the recommendations gathered in the initial two workshops were prioritised and associated with actions and suggested action owners in the third and final workshop. Here it should be clarified that 'priority' is meant as a statement of timeliness more than overall value; in other words, participants were explicitly asked to indicate what should be done the most urgently rather than what should be done versus not done.

As introduced and described above, Figure 2 offers a concise summary of the recommendations as well as the relative priority assigned by different stakeholder groups and a panel of experts. As a first observation the picture shows strong heterogeneity, with different stakeholder groups assigning different priorities to the various recommendations (and occasionally disagreeing amongst themselves). This could be a reflection of the relatively low level of maturity with regards to FAIR data, characterised by many simultaneous challenges, limited information or validation of 'what works', and various actors reviewing or redefining their roles and responsibilities. Still, an area that seems to stand out and confirmed as a priority is that of essential infrastructure

components, including services to automatically create metadata, PID services, and domain-specific ontologies. Complementary to this mostly technical dimension, socially-oriented recommendations around fostering global collaborations and including FAIR in research assessments also scored well across the different stakeholder groups. Suggested actions and action owners for the priority recommendations are collected in Table 1.

The results presented here are naturally a snapshot in time and, as such, represent work in progress. It has proven to be oftentimes challenging to associate more high-level recommendations with pointed, concrete actions and well-defined owners. This workshop series has endeavoured to do that and, while it is hoped that the results presented here will help direct the discussion and spur action, it will no doubt be part of a longer journey with further iterations on the formulation of these recommendations, priorities and actions.

As next steps, the authors were pleased to receive requests to re-use the workshop format to gather and discuss community input in other geographical regions which could help to corroborate findings and paint a fuller, more inclusive picture. In addition, these findings will feed into ongoing work in FAIRsFAIR,[11] the EOSC FAIR Working Group,[12] the Research Data Alliance, OpenAIRE, FREYA,[13] EOSC-hub[14] and other relevant projects. Finally, it is hoped that some readers might recognise themselves as a suggested action owner and find this report helpful to guide them on their path to develop services, infrastructure, tools, ontologies, standards, models, policies and practices that will be supported and valued by the community.

Acknowledgements

---

[11] Notably in Task 2.4 on 'FAIR services & software' as well as the development of recommendations in several work packages. See https://www.fairsfair.eu/the-project
[12] https://www.eoscsecretariat.eu/working-groups/fair-working-group
[13] https://www.project-freya.eu/en
[14] Notably in Task 11.2 Data Management Planning. See https://www.eosc-hub.eu/