Improving the interoperability of biomedical research data

Van Damme, P.

Publication date
2023

Chapter 2

A resource for guiding data stewards to make European rare disease patient registries FAIR

Philip van Damme
Pablo Alarcón Moreno
César H. Bernabé
Alberto Cámara Ballesteros
Clémence M. A. Le Cornec
Bruna Dos Santos Vieira
K. Joeri van der Velde
Shuxin Zhang
Claudio Carta
Ronald Cornet
Peter A.C. 't Hoen
Annika Jacobsen
Morris A. Swertz
Marco Roos
Nirupama Benis

Data Science Journal 19, 1-12 (2023)
DOI: 10.5334/dsj-2023-012

Abstract

Objective

This chapter reports on the development of a dynamic data management planning questionnaire, to guide data stewards of European Reference Network (ERN) rare disease patient registries to make their data Findable, Accessible, Interoperable, and Reusable (FAIR). As part of this work, the questionnaire was validated through expert review and aligned with existing resources on rare diseases and FAIR data management.

Materials and Methods

The questionnaire was developed for the Data Stewardship Wizard, a tool for data management planning. Knowledge sources on FAIR data, ERN patient registries, and data management were used to compose questions. Ten domain experts validated the questionnaire. The topics in the questionnaire were aligned with existing knowledge bases.

Results

A total of 57 questions were included in the questionnaire. Twenty-three references to the FAIR Cookbook and Research Data Management toolkit for Life Sciences were added. Expert validation provided a total of 166 comments on content, structure, or software-related issues. A public instance of the Data Stewardship Wizard was deployed for use by data stewards of ERN patient registries.

Discussion

The questionnaire addresses issues that ERNs encounter when making their registries FAIR and follows the implementation choices made by the European rare disease community. A challenging task for future research is to extend the questionnaire to other types of registries and to validate with users.

Conclusion

This smart questionnaire is the first model created for the Data Stewardship Wizard that helps ERN patient registries with making their data FAIR. It will assist data stewards in aligning their efforts and providing guidance on FAIR data.

Introduction

Up to 36 million people are affected by a rare disease in the European Union (EU), which is around 8% of the total EU population at the time of writing [39]. Like rare disease patients, data about rare diseases are often geographically fragmented. To organize the highly specialized care that patients with a rare disease need, the EU has set up so-called European Reference Networks (ERNs) [40]. By exchanging knowledge and information among healthcare providers, these networks aim to improve access to accurate diagnosis, timely treatment, and appropriate care for people living with rare diseases in Europe. Members of an ERN share expertise on a specific group of diseases (e.g., rare bone or rare kidney diseases). According to the European Medicines Agency, patient registries collect uniform data over time about a population defined by a particular disease, condition, or exposure [41]. A key task of ERNs is setting up and managing patient registries, which are valuable for research, treatment and outcomes monitoring, drug development, and improving quality of care [42]. Standardizing data management practices, allowing for data linking and reuse, has been known to increase the benefits of rare disease patient registries [42–44]. As a result, improving the alignment between ERNs is one of the objectives of the European Joint Programme on Rare Diseases (EJP RD), a project with over 130 institutions from 35 countries, including representatives of all 24 ERNs, designed to establish a self-sustaining infrastructure for rare disease research and care [45]. The EJP RD has been supporting patient registries, managed by ERNs, in making informed choices about their data management, and in harmonizing choices among registries [46].

Wilkinson et al. [15] introduced the Findable, Accessible, Interoperable, and Reusable (FAIR) principles, a set of high-level guidelines for research data management, stating that data should be FAIR for humans and computers. Vieira, Bernabé, and Zhang et al. [46] provided insight into common challenges ERNs encountered when making their patient registry data FAIR, and put forward a list of solutions that may help solve those challenges. To obtain these insights, a team of data stewards specialized in FAIR data have been working closely with ERN patient registries.

Hudson-Vitale and Moulaison-Sandy [47] reviewed scientific research on Data Management Plans (DMPs), and reported that DMPs support data sharing and reuse; however, DMPs were often found to be static documents, making them less effective. Williams et al. [48] presented a framework for DMPs that covers topics such as personnel planning, data elements, data models, software, privacy, and data-sharing practices. Since their introduction, the FAIR principles have been a staple for leveraging DMPs that should produce FAIR data. A tool for assembling DMPs is the Data Stewardship Wizard (DSW) [49]. The DSW uses dynamic questionnaires that provide context-dependent guidance, can generate DMPs from prebuilt templates, and provides metrics for compliance with the FAIR principles. Moreover, the DSW includes an expert-curated knowledge model, which represents a questionnaire, for creating DMPs for life sciences projects. In addition, Mons [50] published a book titled “Data Stewardship for Open Science”, encouraging readers to create their own DMPs using the DSW.

While the default knowledge models of the DSW are successful in helping data stewards create DMPs for projects in the life sciences domain, they fail to cover the domain-specific requirements of ERN rare disease patient registries. In addition, the European rare disease community, represented by the EJP RD, has made implementation choices for the FAIR principles specific to their domain [46]. These choices should be reflected in DMPs of rare disease patient registries. Furthermore, sustaining human support for ERNs is challenging as requirements evolve over time. Hence, there is a strong need for a maintainable data management planning tool tailored to ERN rare disease patient registries.

This chapter reports on the results obtained from (1) creating and validating a data management planning questionnaire to guide ERN patient registries in making their data FAIR, and (2) integration with existing infrastructures around rare diseases and (FAIR) data management.

Materials and Methods

We developed a smart questionnaire for the DSW, i.e., a questionnaire with mostly closed-ended questions that adapts follow-up questions based on previously given answers. This questionnaire — to guide data stewards of ERN patient registries to make their registry data FAIR — was built in four stages: (1) collect and analyze relevant knowledge sources; (2) construct a hierarchical mind map from which a DSW knowledge model was built; (3) acquire feedback from domain experts on the different topics in the questionnaire and DMP, and align the content with existing tools for FAIR data management planning; and (4) set up a public instance of the DSW with the questionnaire preloaded. Figure 2.1 summarizes these stages.

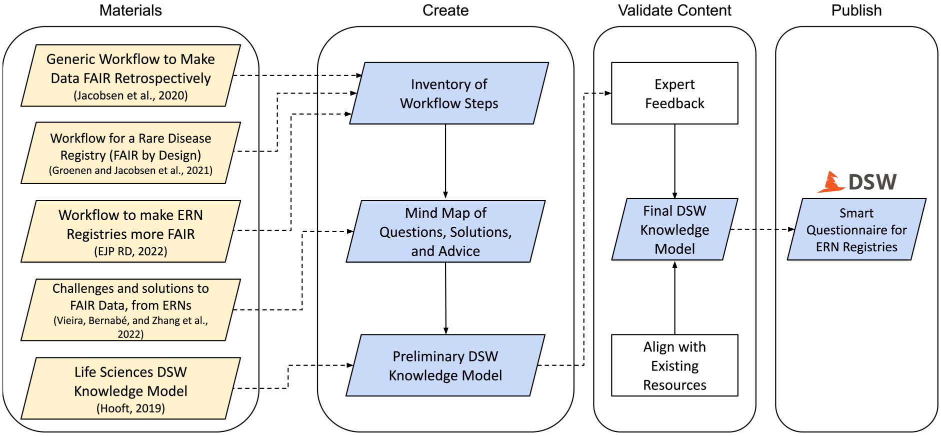

Materials
Create
Validate Content
Publish
Generic Workflow to Make Data FAIR Retrospectively (Jacobsen et al., 2020)
Workflow for a Rare Disease Registry (FAIR by Design) (Groenen and Jacobsen et al., 2021)
Workflow to make ERN Registries more FAIR (EJP RD, 2022)
Challenges and solutions to FAIR Data, from ERNs (Vieira, Bernabé, and Zhang et al., 2022)
Life Sciences DSW Knowledge Model (Hooft, 2019)
Inventory of Workflow Steps
Mind Map of Questions, Solutions, and Advice
Preliminary DSW Knowledge Model
Expert Feedback
Final DSW Knowledge Model
Align with Existing Resources
DSW
Smart Questionnaire for ERN Registries

Figure 2.1: Overview of the stages performed to develop a smart questionnaire for the Data Stewardship Wizard: gathering relevant knowledge sources, developing a Data Stewardship Wizard knowledge model (questionnaire), validating the questionnaire, aligning with the Research Data Management toolkit for Life Sciences (RDMkit, [51]) and FAIR Cookbook [52], and publishing the questionnaire in a Data Stewardship Wizard instance. Abbreviations: Data Stewardship Wizard (DSW); European Reference Networks (ERNs); Findable, Accessible, Interoperable, and Reusable (FAIR). Legend: input (yellow), output (blue).

Materials

First, we gathered relevant sources that provide information or knowledge on making ERN patient registries FAIR. These sources guided our subsequent decisions on the topics and questions to be included in the questionnaire.

- Generic workflow [18]: a step-by-step workflow to make data that has already been collected FAIR.
- Rare disease registry workflow [53]: a workflow designed to make the data of a rare disease patient registry on vascular anomalies FAIR from the moment it is collected.
- Workflow to make ERN registries FAIR [54]: a workflow designed to help ERN patient registries with making their data FAIR.
- Challenges and solutions, from ERN patient registries [46]: an extensive list of 41 challenges and proposed solutions that ERN patient registries encountered when making their data FAIR.
- DSW knowledge model for the life science domain [55]: a knowledge model that includes expert content on data management planning for the life sciences, structured around the research data life cycle.

Create

We created a preliminary knowledge model in three steps. First, we made an inventory of the steps and implementation choices within the three workflows mentioned under "Materials". Second, we built a mind map based on these workflows and ERN challenges, and third, we converted the mind map into a DSW knowledge model. The DSW can export a DMP from a filled-in questionnaire. Questionnaires are generated from knowledge models, which are ordered collections of linked items. A knowledge model contains all the information necessary for generating a questionnaire, such as chapters, questions, descriptions, answer options, and advice bound to answers.

Mind Map

We used the workflows for making data FAIR as the basis for a hierarchical mind map. A mind map was considered an appropriate intermediate step before building a knowledge model because they provide a similar hierarchical structure. This mind map laid the groundwork for what would later become the smart questionnaire. We collaboratively populated the mind map with questions, answer options, and solutions. Solutions were: written advice, software tools, standards, references to internal (i.e., EJP RD) and external resources, or other technical solutions for making data of ERN registries FAIR. Questions and solutions were derived from the work of Vieira, Bernabé, and Zhang et al. [46]. We used MindMeister [56], a cloud-based online mind mapping tool.

Knowledge Model

After completing the mind map, we converted it to a DSW knowledge model. This step is composed of transferring elements from the mind map and adding additional information (such as answer types, descriptions, and titles) using the DSW's built-in knowledge model editor. In addition, to further enrich our model, we reused all seven chapter names and some relevant questions from the life sciences knowledge model of Hooft [55]. These chapters, based on the research data life cycle [49], were found to be a good fit for structuring the content of our mind map. Hence, our knowledge model was built upon the following chapters: administrative information, re-using data, creating and collecting data, processing data, interpreting data, describing data, and giving access to data. Chapters represent sections of a knowledge model. We restructured questions from the mind map to match the chapters when necessary. Additionally, we added tags to questions that addressed a technical implementation choice for findability, accessibility, interoperability, or reusability. Tags are a feature of the DSW that can be used to organize questions, such as to select questionnaire subsets.

Validate Content

To validate the correctness of the content of the questionnaire, we approached ten domain experts and asked for their feedback. Among the invited experts were data scientists, project managers, senior researchers, and software engineers. All experts were affiliated with or involved in the EJP RD. Experts had expertise in authentication and authorization, biobanks, data querying, ERNs, FAIR data, patient consent, privacy legislation, project management, rare diseases, rare disease patient registries, record linkage, semantic models, and software architecture. Experts were asked to only appraise content relevant to their expertise. For example, an expert on patient consent would be asked to review all questions related to consent. Experts reviewed individual questions, the structure of the knowledge model, and additional information presented along with the questions and answers. Feedback was collected through a spreadsheet form, via video call, or both. We curated the received expert reviews to remove duplicate comments and to clarify what changes should be made to the knowledge model. We divided the curated feedback into three categories: textual change (question, answer (option), or description), structural change (e.g., change the question order), or software issue. We then updated the knowledge model according to the feedback.

Finally, we aligned the questionnaire with two existing resources that offer a plenitude of knowledge on how to make data FAIR. That is, the Research Data Management Toolkit for Life Sciences (RDMkit) and the FAIR Cookbook [51,52]. We added references to pages from the RDMkit or recipes from the FAIR Cookbook to any description or advice in the questionnaire that mentioned a topic also covered by one or both resources.

Publish

Publishing involved hosting a public instance of the DSW with our knowledge model preloaded. We hosted this instance on the servers of ELIXIR's Czech Republic Node, which also manages support and operation of the DSW [57]. Existing privacy policies apply to this instance. The knowledge model source files were made available on a public repository [1].

Results

The inventory of workflow steps to make data FAIR and implementation choices suggested by the EJP RD comprises nine steps, 19 topics related to those steps, and 12 implementation choices (e.g., a certain tool or standard). Table 2.1 shows an overview of this inventory.

Table 2.1: Overview of the workflow steps and inventory of topics and implementations.

Workflow Step	Related Topics	Implementation
1. Identify FAIR objectives and expertise		
	a. Defining objectives	
	b. Giving training	
	c. Hiring of personnel	
2. Define data elements to be collected		
	a. Common data elements	CDE core elements [58]
	b. Data dictionary	
	c. Central metadata repository registration	ERDRI. mdr [59]
3. Define metadata elements to be collected		
	a. Machine interpretable metadata	EJP RD metadata model [60]
	b. Metadata store	FAIR Data Point [61]
4. Create a semantic data model		
	a. Reuse of existing model(s)	CDE semantic model [62] CDISC ODM [63] HL7 FHIR [38] OMOP CDM [64]
5. Obtain consent		
	a. Standardized informed consent form	ERN ICF [65]
6. Enter (FAIR) data		
	a. Electronic Data Capture systems	
7. Standardize metadata		
	a. Metadata model(s)	EJP RD metadata model [60]
	b. Standard terminology	CDE semantic model terminology [66]
8. Transform (meta)data to RDF		
	a. Data transformation	CDE in a box [67]
	b. Terminology mappings	
9. Manage authentication and authorization		
	a. Authorization roles	
	b. Access conditions	
	c. Data pseudonymization	
	d. Querying	

Abbreviations: Common Data Elements on rare disease registration (CDE), European Platform on Rare Disease Registration MetaData Repository (ERDRI.mdr), European Reference Network (ERN), Health Level 7 Fast Healthcare Interoperability Resources (HL7 FHIR), Clinical Data Interchange Standards Consortium Operational Data Model (CDISC ODM), Observational Medical Outcomes Partnership Common Data Model (OMOP CDM), Findable, Accessible, Interoperable, and Reusable (FAIR), Informed Consent Form (ICF), Resource Description Framework (RDF).

Steps, topics, and implementations were translated into questions, answers, and advice. For example, the topic "defining objectives" was rephrased as "Have you defined objectives?". Eventually, the mind mapping process resulted in 22 out of 41 ERN challenges identified by Vieira, Bernabé, and Zhang et al. [46] to be included as a question, answer, advice, or a combination of the three. Challenges that were categorized under "community", i.e., alignment between ERNs, were not included as questions but were indirectly addressed by using the DSW.

[1] A user guide and the knowledge model source files are available at https://github.com/ejp-rd-vp/smart-guidance.

That is to say, enabling ERN data stewards to use the DSW with our questionnaire untangles those challenges in part. For example, ERNs found that they were unaware of choices other ERNs made, which they can now share through the DSW. Similarly, one challenge categorized under "training" was also indirectly addressed (need for more information on activities of the EJP RD). Table 2.2 shows the number of challenges included in the questionnaire and the motivation for why some were not. The full list of challenges can be found in the original publication [46].

Table 2.2: Challenges from [46] that were included in the questionnaire during the mind mapping phase. Challenges marked as indirectly covered are not specifically mentioned in the questionnaire but were solved solely by the use of the Data Stewardship Wizard and the questionnaire. Categories originate from [46].

Category	**Directly Included**	**Indirectly Included**	**Motivation**
Community	0 out of 7	7 out of 7	All challenges addressed a lack of alignment between registries, the DSW questionnaire solves this issue.
Implementation	7 out of 9	0 out of 9	Two not-included challenges were irrelevant at the time of developing the questionnaire.
Legal	3 out of 5	0 out of 5	Two not-included challenges addressed a tool that was not relevant for developing the questionnaire.
Modeling	3 out of 5	0 out of 5	Two not-included challenges addressed issues that were too specific.
Training	9 out of 15	1 out of 15	Five not-included challenges addressed irrelevant tools. One indirectly covered challenge was not mentioned specifically in the questionnaire but could be deducted from the information.
All Categories	**22 out of 41**	**8 out of 41**	

The mind map was converted into a preliminary DSW knowledge model. That is, questions, answers, and advice were added to the knowledge model based on the mind map. We reused one question from the life sciences knowledge model of Hooft [55]: "Who is a contributor to the DMP". Once this preliminary version of the questionnaire was available in the DSW, we started the validation process.

Validating the correctness of the content of the questionnaire resulted in an updated version of our knowledge model. A total of 10 experts reviewed the content of the questionnaire. We received a total of 166 comments. Each chapter was assigned at least seven experts, all experts reviewed the questions in "Processing data" and "Interpreting data". Table 2.3 shows an overview of the comments per chapter and category. Duplicate comments often regarded textual issues on flow or clarity. Experts also provided references to additional resources, such as web pages with more information on a certain topic. Structural changes asked for moving a question up or down the hierarchy or to another chapter. Software issues were related to issues with using the DSW interface, such as a non-functional button or a page that would not load. These issues were solved by updating to the latest version of the DSW.

After processing the feedback and updating the knowledge model, the questionnaire has 57 questions. A total of six questions are open-ended and 51 questions are closed-ended. In total, 10 references were added to recipes in the FAIR Cookbook and 13 references to pages of the RDMkit. Three questions were tagged as an implementation choice for findability, six to accessibility, 14 to interoperability, and 21 to reusability. Thirteen questions were not tagged because they did not cover implementation choices but rather aspects like training, objectives, or administrative topics. Table 2.4 shows the number of questions and external references per chapter.

Table 2.3: Quantification of the received feedback per chapter. Feedback is categorized as textual change, structural change, or software issue.

Chapter	Textual Changes	Structural Changes	Software Issues
Administrative information	18	5	3
Re-using data	13	2	0
Creating and collecting data	3	2	2
Processing data	19	3	0
Interpreting data	47	8	0
Describing data	10	1	0
Giving access to data	27	3	0
All Chapters	**137**	**24**	**5**

Table 2.4: Questions and external references per chapter. Top-level questions are questions that precede all other questions and are always presented to a user.

Chapter	Top-Level Questions	Total Questions	References to FAIR Cookbook	References to RDMkit
Administrative information	6	15	1	4
Re-using data	2	9	3	3
Creating and collecting data	2	5	1	1
Processing data	1	5	0	2
Interpreting data	2	12	4	1
Describing data	2	4	0	0
Giving access to data	4	7	1	2
All Chapters	**19**	**57**	**10**	**13**

Abbreviations: Findable, Accessible, Interoperable, Reusable (FAIR), Research Data Management toolkit for Life Sciences (RDMkit, [51]). FAIR Cookbook by FAIRplus [52].

Figure 2.2 depicts a simplified view of our knowledge model and includes all topics covered by the questionnaire. This is the final model that was constructed after expert validation. The questionnaire covers a broad range of topics: building and training a team of professionals, defining data management objectives, (meta)data modeling, data elements, using common standards, using common terminology, data pseudonymization, electronic data capture, querying, metadata exposure, authentication and authorization, and informed consent. Figure 2.3 provides a screenshot of the chapters and top-level questions. Figure 2.4 provides a screenshot of how the questionnaire is presented to a user.

Discussion

The purpose of this work was to develop a smart questionnaire that guides data stewards working to make data of ERN rare disease patient registries FAIR. Data stewards of patient registries will increasingly have to manage data in ways that comply with implementation choices of the FAIR principles as recommended by their community. Standardizing data management practices of patient registries enables the virtual pooling of otherwise sparse and geographically scattered rare disease data, increasing their usefulness for effective research and care. However, standardization for each of the FAIR principles in this domain is complex. ERNs face the challenge of registering data of thousands of diseases from many different sources and making that data as usable as possible within a global health data ecosys-

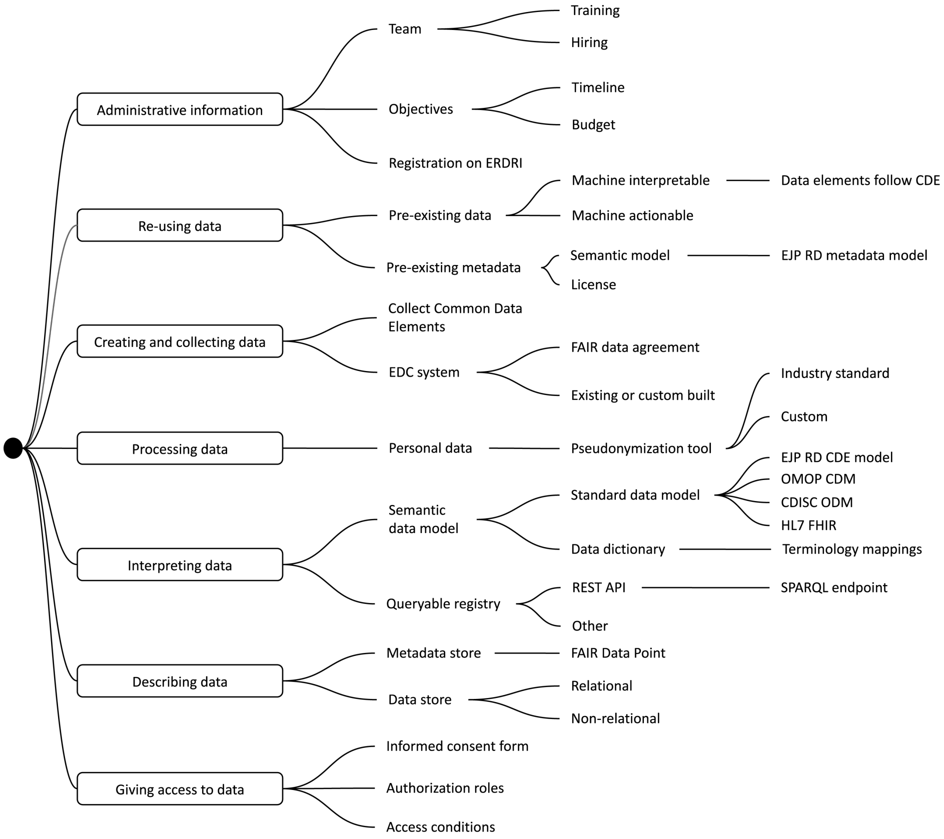
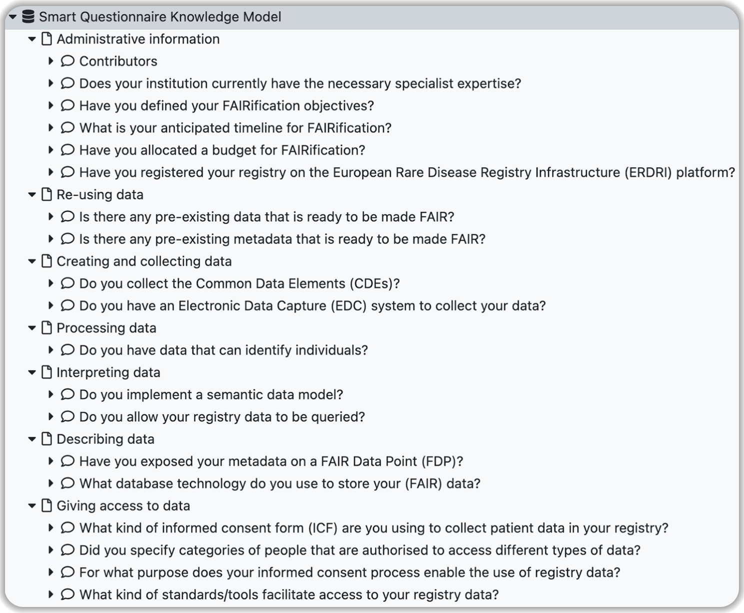
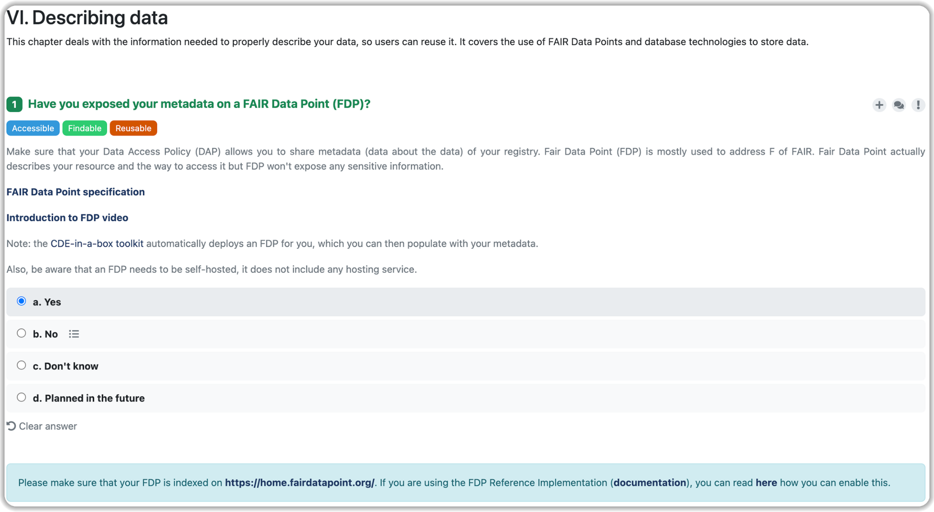


Figure 2.2: Simplified view of the knowledge model. Abbreviations: Common Data Elements (CDE), Electronic Data Capture (EDC), European Joint Programme on Rare Diseases (EJP RD), European Platform on Rare Disease Registration (ERDRI), Findable, Accessible, Interoperable, and Reusable (FAIR), Health Level 7 Fast Healthcare Interoperability Resources (HL7 FHIR), Clinical Data Interchange Standards Consortium Operational Data Model (CDISC ODM), Observational Medical Outcomes Partnership Common Data Model (OMOP CDM), REpresentational State Transfer Application Programming Interface (REST API), SPARQL Protocol and RDF (Resource Description Framework) Query Language (SPARQL).



- Smart Questionnaire Knowledge Model
  - Administrative information
    - Contributors
    - Does your institution currently have the necessary specialist expertise?
    - Have you defined your FAIRification objectives?
    - What is your anticipated timeline for FAIRification?
    - Have you allocated a budget for FAIRification?
    - Have you registered your registry on the European Rare Disease Registry Infrastructure (ERDRI) platform?
  - Re-using data
    - Is there any pre-existing data that is ready to be made FAIR?
    - Is there any pre-existing metadata that is ready to be made FAIR?
  - Creating and collecting data
    - Do you collect the Common Data Elements (CDEs)?
    - Do you have an Electronic Data Capture (EDC) system to collect your data?
  - Processing data
    - Do you have data that can identify individuals?
  - Interpreting data
    - Do you implement a semantic data model?
    - Do you allow your registry data to be queried?
  - Describing data
    - Have you exposed your metadata on a FAIR Data Point (FDP)?
    - What database technology do you use to store your (FAIR) data?
  - Giving access to data
    - What kind of informed consent form (ICF) are you using to collect patient data in your registry?
    - Did you specify categories of people that are authorised to access different types of data?
    - For what purpose does your informed consent process enable the use of registry data?
    - What kind of standards/tools facilitate access to your registry data?

Figure 2.3: Screenshot of the knowledge model with top-level questions (Data Stewardship Wizard knowledge model editor module).



VI. Describing data

This chapter deals with the information needed to properly describe your data, so users can reuse it. It covers the use of FAIR Data Points and database technologies to store data.

1 Have you exposed your metadata on a FAIR Data Point (FDP)?

Make sure that your Data Access Policy (DAP) allows you to share metadata (data about the data) of your registry. Fair Data Point (FDP) is mostly used to address F of FAIR. Fair Data Point actually describes your resource and the way to access it but FDP won't expose any sensitive information.

FAIR Data Point specification

Introduction to FDP video

Note: the CDE-in-a-box toolkit automatically deploys an FDP for you, which you can then populate with your metadata.

Also, be aware that an FDP needs to be self-hosted, it does not include any hosting service.

● a. Yes

○ b. No

○ c. Don't know

○ d. Planned in the future

Clear answer

Please make sure that your FDP is indexed on https://home.fairdatapoint.org/. If you are using the FDP Reference Implementation (documentation), you can read here how you can enable this.

Figure 2.4: Screenshot of the first question of the “Describing data” chapter (Data Stewardship Wizard questionnaire module).

tem. Our questionnaire addresses those challenges ERNs were known to commonly face and provides guidance according to the FAIR infrastructure set up by the EJP RD. The questionnaire acts as a checklist for making rare disease registries FAIR: data stewards can make sure that all boxes are checked.

Our questionnaire covers the process of making data in patient registries FAIR. Although the questions and advice are made for ERN patient registries, the questionnaire content was based on prior knowledge of FAIR data and experiences in the European rare disease community. For instance, annual "bring your own data" workshops have brought together FAIR data experts and rare disease data managers since 2014 [68]. Workshops such as these have provided a valuable source of challenges regarding the FAIR guiding principles and proposed implementations thereof (e.g., see the work of Jacobsen et al. [16]). Furthermore, they affirmed our motivation for designing a smart questionnaire that enables data stewards to begin their FAIR journey from a variety of starting points. As a result of tailoring each topic and question to the unique needs of ERN patient registries, we have filled the gap in having a data management tool that is suitable for rare disease registries in Europe. Since this work is part of ongoing efforts of the EJP RD, integration of the DSW and questionnaire with the European infrastructure for rare disease research will be a natural next step.

Previous studies focused on DMP requirements [48], and DMPs for the life sciences domain [49, 55]. Williams et al. concluded that while most DMPs included components describing data reuse and sharing, few DMPs described data collection and processing practices. These last two are particularly hard to fix, as the quality of poorly collected data can most likely not be improved in retrospect. We were able to address all four topics in our questionnaire. Creating DMPs for projects in the life sciences was addressed by the original authors of the DSW. We found that, by reusing parts of their knowledge model, we were able to structure our questionnaire according to a well-established model. Moreover, Jones et al. [69] concluded that DMPs are essential for FAIR data stewardship. By adopting the DSW as a tool for making ERN patient registries FAIR, we believe our work aligns with that conclusion.

Our work has some limitations. We validated the content of the questionnaire for correctness through expert feedback, but we did not validate the impact of the questionnaire on its intended users. Therefore, further research is needed to determine whether ERN registry data stewards benefit from our tool. Furthermore, the questions and advice are specific to the situation of ERN patient registries and cannot be extrapolated to other registries or projects without modifications. Our work mainly focused on guiding ERN patient registries in making their data FAIR; nevertheless, there is clear value in aligning more types of registries as well. Registry types outside of rare diseases, as well as non-European rare disease patient registries, could fall into this category.

Our work also has several strengths. First, navigating through FAIR implementation choices via questions and answers is a different experience from filling out DMP checklists. It is anticipated that this will lead to an increase in the quality of DMPs. Secondly, through the DSW, data stewards can learn from the implementation choices of others. Thus, it complements in-person training and contributes to community convergence. Thirdly, we created a single place where ERN data stewards can go for guidance on making their registry FAIR. This makes maintaining and updating the knowledge in the questionnaire easier compared to having various sources in different locations. The knowledge model can be improved by learning from users who will fill out questionnaires on the DSW platform. Moreover, the DSW software is being actively maintained, and hosting our instance on the ELIXIR infrastructure means that it can be sustained beyond the lifetime of the EJP RD.

The knowledge model we developed is publicly available[2] and can be used by others to build upon or to reuse parts from. For exporting the DMP, we use a default DSW template, which we intend to customize in the near future. It may be possible to improve the guidance offered to ERN data stewards through further customization of this template. Additional research is needed to quantify the impact of our questionnaire on the (process leading to) "FAIRness" of ERN patient registries. Another challenging task for further research is to extend the questionnaire to other types of resources by collaborating with resource owners and users.

Conclusions

The developed smart questionnaire for the DSW is a promising method for guiding data stewards in making their registry data FAIR. It is the first model created for the DSW that helps to standardize data management practices among ERN patient registries. Future research should focus on user validation and extending the questionnaire beyond the realm of ERNs.

[2] The smart questionnaire is available at https://smartguidance.ejprarediseases.org (registration is required).