

Building expertise on FAIR through evolving Bring Your Own Data (BYOD) workshops: describing the data, software, and management focused approaches and their evolution

César H. Bernabé^{1‡}, Lieze Thielemans^{2‡}, Rajaram Kaliyaperumal^{1‡}, Claudio Carta^{3‡}, Shuxin Zhang^{4,5}, Celia W.G. van Gelder⁶, Nirupama Benis^{4,5}, Luiz Olavo Bonino da Silva Santos^{1,7}, Ronald Cornet^{4,5}, Bruna dos Santos Vieira^{8,9}, Nawel Lalout^{9,10}, Ines Henriques⁹, Alberto Cámara Ballesteros¹¹, Kees Burger¹, Martijn G. Kersloot^{4,5,12}, Friederike Ehrhart¹³, Esther van Enckevort¹⁴, Chris T. Evelo^{13,15}, Alasdair J. G. Gray¹⁶, Marc Hanauer¹⁷, Kristina Hettne¹⁸, Joep de Ligt¹⁹, Arnaldo Pereira²⁰, Núria Queralt-Rosinach¹, Erik Schultes^{21,22}, Domenica Taruscio³, Andra Waagmeester²³, Mark D. Wilkinson¹¹, Egon L. Willighagen¹³, Mascha Jansen⁶, Barend Mons^{1,21}, Marco Roos^{1*□}, Annika Jacobsen^{1*□}

- 1** Biosemantics Group, Department of Human Genetics, Leiden University Medical Centre, Leiden, The Netherlands
- 2** Department of Ageing & Health, Guy's and St Thomas' NHS Trust, UK
- 3** National Centre for Rare Diseases, Istituto Superiore di Sanità, Rome, Italy
- 4** Department of Medical Informatics, Amsterdam UMC location University of Amsterdam, Meibergdreef 9, Amsterdam, The Netherlands
- 5** Amsterdam Public Health, Methodology & Digital Health, Amsterdam, The Netherlands
- 6** DTL Dutch Techcentre for Life Sciences, Utrecht, The Netherlands
- 7** Semantics, Services and Cybersecurity group, University of Twente, Enschede, The Netherlands
- 8** Department of Medical Imaging, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands
- 9** Department of Medical BioSciences, Radboud University Medical Center, Nijmegen, The Netherlands
- 10** The Duchenne Parent Project, Veneendaal, The Netherlands
- 11** Departamento de Biotecnología-Biología Vegetal, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Centro de Biotecnología y Genómica de Plantas. Universidad Politécnica de Madrid (UPM) - Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria-CSIC (INIA-CSIC). Campus Montegancedo, Pozuelo de Alarcón (Madrid), Spain.
- 12** Castor, Amsterdam, The Netherlands
- 13** Department of Bioinformatics - BiGCaT, NUTRIM/MHeNs, Maastricht University. Maastricht, The Netherlands
- 14** University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
- 15** Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, The Netherlands
- 16** Department of Computer Science, Heriot-Watt University, Edinburgh, UK
- 17** INSERM, US14—Orphanet, Plateforme Maladies Rares, Paris, France
- 18** Centre for Digital Scholarship, Leiden University Libraries, The Netherlands
- 19** Center for Molecular Medicine, Cancer Genomics Netherlands, Department of Genetics, University Medical Center Utrecht, University Utrecht, The Netherlands
- 20** University of Aveiro, DETI/IEETA, Aveiro, Portugal

21 GO FAIR Foundation, Leiden, The Netherlands

22 The Leiden Academic Center for Drug Research, Leiden, The Netherlands

23 Micelio, Ekeren-Antwerp, Belgium

‡These authors share first authorship

☒These authors share last authorship

* m.roos@lumc.nl, a.jacobsen@lumc.nl

Abstract

Since 2014, “Bring Your Own Data” workshops (BYODs) have been organised to inform people about the process and benefits of making resources Findable, Accessible, Interoperable, and Reusable (FAIR, and the FAIRification process). The BYOD workshops’ content and format differ depending on their goal, context, and the background and needs of participants. Data-focused BYODs educate domain experts on how to make their data FAIR to find new answers to research questions. Management-focused BYODs promote the benefits of making data FAIR and instruct project managers and policy-makers on the characteristics of FAIRification projects. Software-focused BYODs gather software developers and experts on FAIR to implement or improve software resources that are used to support FAIRification. Overall, these BYODs intend to foster collaboration between different types of stakeholders involved in data management, curation, and reuse (e.g. domain experts, trainers, developers, data owners, data analysts, FAIR experts). The BYODs also serve as an opportunity to learn what kind of support for FAIRification is needed from different communities and to develop teaching materials based on practical examples and experience. In this paper, we detail the three different structures of the BYODs and describe examples of early BYODs related to plant breeding data, and rare disease registries and biobanks, which have shaped the structure of the workshops. We discuss the latest insights into making BYODs more productive by leveraging our almost ten years of training experience in these workshops, including successes and encountered challenges. Finally, we examine how the participants’ feedback has motivated the research on FAIR, including the development of workflows and software.

Introduction

The FAIR data principles address critical factors to make the analysis of multiple sources more efficient by improving their Findability, Accessibility, Interoperability, and Reusability for humans and computers [1]. The process of making data FAIR (“FAIRification”), although partially supported by software (e.g. data transformation tools) and standards (e.g. ontologies), relies on expert knowledge about the data generating domain (domain experts and data owners) and about FAIR-related aspects such as metadata design, conceptual modelling, licensing definition, and use of identifiers (FAIR experts). Given the high demand to make resources FAIR and a shortage of FAIR expertise, a series of “Bring Your Own Data” workshops (BYODs) have been organised since 2014 to bring together expert knowledge to accelerate the practical FAIRification of resources (e.g. datasets, registry information, ontologies). The bidirectional learning experience between domain and FAIR experts results in making the BYODs a mutually beneficial experience. Attendees (domain experts) receive hands-on guidance in making their data FAIR, while trainers (FAIR experts) gain valuable insights to improve their own training skills, topics, and materials, and develop more effective FAIR support tools, processes, and guidelines.

The first BYOD workshop took place six months after the “Jointly Designing a Data FAIRPORT” workshop [2], which marked the inception of the FAIR principles. Subsequently, the first BYOD for rare disease registries and biobanks, held in November 2014, initiated an annual series of BYODs for this specific domain. Initially, data-focused BYODs emerged from the need to train people on FAIRification and therefore focused on making data resources FAIR. Over time, as the FAIR community matured, it became clear that different types of BYODs were necessary to meet different contexts, needs and backgrounds of attendees. Consequently, two additional types of BYOD structures were designed: management- and software-focused BYODs. The former aims at informing managers and policy-makers on the added benefits of FAIR and the requirements for FAIRification. The latter focus on developing software tools that support the process of making data FAIR, or tools and standards that enhance the FAIR level of resources.

The remainder of this paper is organised into four sections. The next section describes the three types of BYODs. The section “the evolution of BYODs” lists the BYODs organised since 2014 and reports on the first BYOD workshop and the series of BYODs for rare disease registries, with emphasis on how these workshops have led to the improvement of the content and didactical aspects of the BYODs. Then, we discuss the impact of the BYODs on the FAIR community and on the expert domains (i.e. plant breeding, and rare disease registries and biobanks). In this paper, we mention different types of experts. For clarity, we refer to “X expert” as an expert specialised in a certain tool, standard, or knowledge. For instance, “FAIR expert” refers to an expert with knowledge and experience in FAIR.

The three types of BYOD structure

Despite catering for different contexts and types of participants, all BYODs share the same overarching goal of fostering expertise in FAIR-related topics while cultivating community confidence in the benefits of having FAIR resources. Table 1 summarises the main aspects of the three different BYOD structures, which are further described in the subsections that follow.

In addition to the two- or three-day duration of the BYODs, some also contain preparatory and follow-up phases for attendees and trainers. Attendees are invited to participate in introductory webinars, to prepare for the workshop, and post-BYOD follow-up meetings, for support on subsequent activities. The introductory webinars aim to familiarise attendees with FAIR and initial FAIRification needs (e.g. identification of goals and required domain expertise). In the post-BYOD follow-up meetings, attendees are advised on other FAIRification challenges that might appear after the BYOD. The preparatory and follow-up phases provide opportunities for trainers to prepare and evaluate the workshop agenda, training materials, and methods of instruction. Follow-up phases are also used by trainers to plan for improvements in future editions of the workshop in response to feedback from participants.

Data-focused BYOD workshops

During the data-focused BYODs, attendees are divided into groups, with at least one trainer allocated per group. Each group can use their own data or request synthetic data. Collaboratively, the groups transform their data into FAIR data by following a step-by-step FAIRification process.

In most workshops, we have followed a FAIRification workflow where data is made FAIR retrospectively (after data collection - *post hoc*) and semi-automatically (see Figure 1). This workflow was developed based on emerging FAIRification steps observed

Table 1. An overview of the main characteristics of the BYODs structure.

	Data	Management	Software
<i>Required technical knowledge for attendees</i>	Intermediate	Low	High
<i>Learning Format</i>	- Knowledge exchange - Hackathon - Lectures	- Knowledge exchange - Lectures	- Knowledge exchange - Hackathon
<i>Main goals</i>	- Answer research questions using FAIR data - Make resources FAIR - Train domain experts on FAIR(ification)	- Inform participants about the added benefits of FAIR - Inform managers on the characteristics of FAIRification projects	- Develop software that support enabling FAIR on resources (e.g. FAIR Data Point) - Develop resources that support the FAIRification process (e.g. data transformation tools)
<i>Main tasks</i>	- Semantic modelling of (meta)data - Hosting and querying of FAIR data	- Hands-on scenarios simulating the benefit of FAIR resources - Plenary sessions to discuss the characteristics of FAIRification projects	- Solution brainstorming - Solution implementation - Testing of implemented solution
<i>Profile of attendees</i>	- Researchers - Domain experts	- Project managers - Registry managers - Patient representatives - Policy makers - Funders	- Researchers on FAIR - Software Developers - FAIR data stewards - Domain experts
<i>Profile of trainers</i>	- Experts in FAIR - FAIR data stewards - Ontologists - Standards specialists	- FAIRification project managers - FAIR data stewards - Decision/policy makers with knowledge on FAIR	- Researchers on FAIR - Developers - FAIR data stewards - Domain experts
<i>Commonly used tools and standards</i>	- Ontologies and metadata models - FAIR Data Point - Domain specific standards (e.g. CDE Semantic Model [ref]) - FAIR enabling standards (e.g. DCAT [40])	- FAIRification workflows - Collaborative brainstorming tools (e.g. mind maps, black boards)	- Collaborative brainstorming tools - Software development resources (e.g. programming languages such as Python)
<i>Expected Outputs</i>	- FAIR resource - Answer to research question(s)	Audience knowledgeable about: - the characteristics of FAIRification - the added benefits of FAIR	- FAIR enabling resource - FAIRification supporting resource

The first row describes the technical knowledge (e.g. on data transformation algorithms) required from the attendees. The second row lists the learning formats that can be used in the BYODs. The third and fourth rows depict the main goals and tasks of the BYODs, respectively. The fifth and sixth rows show the expected backgrounds of the attendees and trainers of the BYODs, respectively. Note that project managers refer to those who manage projects that can benefit from FAIR(ification), such as data collection projects, while standard specialists refer to experts in any standard that enables FAIR (e.g. FAIR Data Point [3]). The seventh row lists the standards and tools commonly used in the FAIRification, as well as development tasks or lectures given during the BYODs. The last row describes the expected outputs for each BYOD format.

in early BYODs. It should be noted, however, that previous BYODs may have deviated from this structure while evolving towards the current format. Additionally, more recent workshops focused on making data FAIR by design (automatically during data

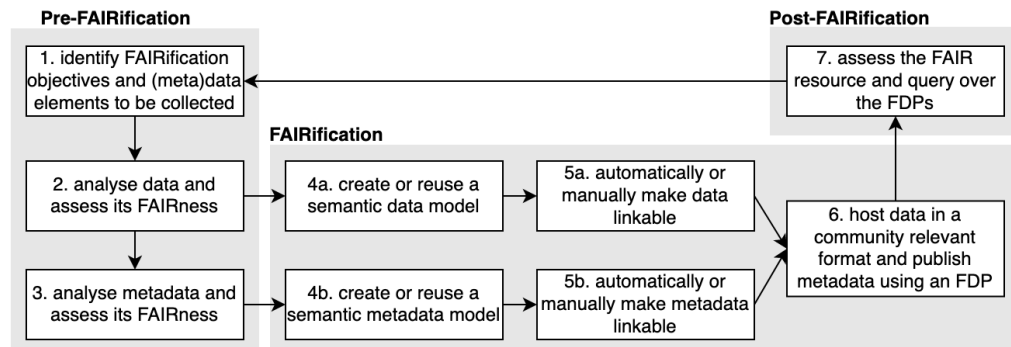


Fig 1. Illustration of the FAIRification workflow used in the data-focused BYODs, modified from [4] and [5]

collection - *de novo*) (e.g. [5]). Figure 1, which is adapted from [4] and [5], illustrates the FAIRification workflow used in recent BYODs. The workflow is divided into three phases: 1) Pre-FAIRification, 2) FAIRification, and 3) Post-FAIRification, which are each subdivided into clear steps. These hands-on phases are usually accompanied by lectures about FAIR related topics and plenary sessions where participants can share their experiences with FAIRification, including challenges and success cases (as listed in Table 1).

The Pre-FAIRification phase is composed of three steps. In step 1, to identify FAIRification objectives and (meta)data elements to be collected, the groups define driving objectives and research question(s) focusing on using their sample data in combination with other FAIR data. Next, following steps 2 and 3, the groups closely investigate the representation (syntax) and meaning (semantics) of their data and the metadata (i.e. description of data). In our experience, metadata such as the (meta)data’s licence and provenance information is often not available *a priori* and therefore needs to be gathered during the BYOD workshop. Finally, before doing any actual FAIRification, the FAIR status of the data is assessed by using tools such as the FAIR Evaluation Services [6] (step 3) (see [7] for other FAIR assessment services).

In the FAIRification phase, the groups create or reuse a conceptual model to describe the data elements and their relationship (step 4a), and a metadata model to provide information about the data (step 4b). These conceptual models must contain, at a minimum, the data elements required to answer their driving research question(s). These conceptual models are semantically enriched by binding the models’ concepts to terms from reference ontologies. In step 5a, the data is made machine-readable (i.e. in a format that can be processed by a computer) by using the semantic data model and existing tooling to generate an ontologised version of the data manually (e.g. FAIRifier [8]) or automatically (e.g. Castor EDC [9], MOLGENIS [10]). The metadata is also made machine-readable (step 5b) by using metadata standards (e.g. Data Catalogue Vocabulary (DCAT) [11]). Finally, the machine-readable metadata is made available using the FAIR Data Point (FDP) (step 6) [3] and the machine-readable data is hosted using a community relevant file format (e.g. RDF [12]).

In the post-FAIRification phase, the driving research question(s) defined in step 1 are answered using the newly created FAIR data (step 7). Here, the FAIR status of the data resource is reassessed and compared to the assessment done in the pre-FAIRification phase to verify if the improvement of the FAIR level of the data meets the goals previously defined.

To illustrate, in the pre-FAIRification phase, a group defines “finding new treatment candidates for untreated rare disease patients” as a driving goal, and reuses data from

different rare diseases registries to achieve this goal. The data includes information on diagnosis, symptoms and treatments, and metadata includes information such as the (meta)data's licence (e.g. CC BY-NC 4.0 [13]) and provenance (e.g. from which registry the data originates). In the FAIRification phase, the group adopts the CDE Semantic Model [14] (step 4a) as the semantic data model and the EJP RD metadata model [15] as the semantic metadata model (step 4b). Reusing the ontologies from the models adopted in steps 4a and 4b supports making the data and metadata linkable (steps 5a and 5b). Finally, the newly linked (meta)data is hosted and published using a FDP (step 6). During the post-FAIRification phase, the group leverages the FAIR data they have created to address their research question by writing federated queries. For instance, they may query their FAIR data with other public resources to identify treatment candidates for patients with similar symptoms.

Management-focused BYOD workshops

The management-focused BYOD workshops are geared towards informing registry and project managers, patient representatives, and decision-makers about the characteristics of FAIRification, including the associated costs, time, expertise, and effort required. The need for this type of BYOD emerged due to the growing adoption of FAIR in various institutions, which has required personnel in high-level positions to become familiar with the benefits and prerequisites of data FAIRification. As a result, these workshops place less emphasis on technical work and more on general considerations of FAIRification. The management-focused BYOD is divided into three phases: (i) understanding the problem of not having FAIR data, (ii) acquiring knowledge about FAIR and FAIRification, and (iii) training on FAIRification project management.

The first phase of a management-focused BYOD is executed in an interactive manner, typically through the use of simulated case scenarios that recreate the challenge of dealing with incomprehensible and non-interoperable data. To highlight the importance of FAIR data, attendees are tasked with challenges that require connecting data from multiple sources, while being presented with non-standardised and multilingually annotated data in different formats, making the task more difficult to accomplish.

In the second phase, attendees learn about the benefits of FAIR and the main steps of FAIRification. The learned benefits aim to address the challenges identified in phase one. Plenary and hands-on sessions provide practical experience in FAIRification related tasks, including conceptual modelling, making metadata findable, and using ontologies and FAIR compliant Electronic Data Capture (EDC) systems. This phase is typically concluded by revisiting the mock case from the first phase, but this time using FAIR data, thus demonstrating how the previously identified challenges can now be solved more easily and efficiently.

In the third phase, participants discuss the implications (e.g. budget, time, required expertise and infrastructure) of FAIR for project managers and policymakers. After the plenary sessions, attendees have a hands-on session on how to create their FAIRification team.

A real-world example of this structure can be visualised on the agenda of recent management-focused BYODs organised for rare disease registries and biobanks (e.g. [16]). For instance, in the one held online in 2022 [16], attendees experienced the problems of not having FAIR data through a digital game where they had to find treatments for new patients in different datasets organised in a non standardised manner (e.g. using synonyms for equivalent concepts) and presented in several languages (e.g. Mandarin, Dutch, and Spanish). Thereafter, lectures and discussion sessions on topics such as FAIRification steps, conceptual modelling, ontologies and querying informed the attendees about FAIR-related aspects. On the second day, the

attendees played the same digital game, only this time with FAIR data, which allowed them to accomplish the goal of finding treatment in distributed datasets. After lectures and discussions on the benefits of FAIR, participants exchanged experiences about the implications of data FAIRification for registry managers.

Software-focused BYOD workshops

The main goal of software-focused BYOD workshops is to create software that supports FAIRification, or software and standards that increase the FAIR level of resources, as shown in Table 1. Participants of software-focused BYODs include researchers working on FAIR-related projects, FAIR data stewards [17], developers and, in certain cases, domain experts. In this type of workshop, trainers and attendees come from similar backgrounds, working together to exchange knowledge while tackling the same goal. This type of BYOD is organised in a hackathon setting with five phases:

1. Understanding the problem: participants discuss the need or problem to be solved (e.g. facilitating metadata publication)
2. Proposing solutions: participants are invited to brainstorm solutions (e.g. using brainstorming tools such as mind maps) to the problem described in the previous phase (e.g. developing software to support the creation and publication of FAIR metadata)
3. Prioritising tasks: the implementation tasks are ordered by importance, and then selected for implementation (e.g. developing a proof-of-concept software that creates machine-readable metadata from an Excel sheet and publishes it via an FDP)
4. Coding and experimenting: the prioritised tasks are implemented and the resulting implementation is tested (e.g. implementing and testing the proof-of-concept software with mock data).
5. Reporting: the implementation is reported and published (e.g. a paper or website documenting the script developed during the hackathon)

Most software-focused BYODs are typically structured around iterative cycles. After a set period of time, participants convene to report on their group status and the findings from their tasks. They can then decide to switch or merge groups, get advice from others and/or continue their tasks. As a result, the agenda of the software-focused BYOD is adaptable to the requirements that emerge during the workshop. At the end of the BYOD, conclusions on which solutions to follow up are made. Adaptations of current tools, prototypes, proof-of-concept implementations or architectural designs are examples of outcomes of software-focused BYODs.

The “hackathon to make MOLGENIS FAIR” [18], which took place in 2016, is a real-world example of a software-focused BYOD. During this event, software developers and FAIR experts worked collaboratively to create a proof-of-concept of making MOLGENIS FAIR. MOLGENIS [10] is an open-source data platform for the management of scientific data. By the end of the hackathon, the team had implemented an application programming interface (API) to publish datasets in MOLGENIS as FDPs.

The evolution of BYODs

Table 2 highlights the BYODs held from 2014 to 2023. For context, the table includes the date of the “Jointly Designing a Data FAIRPORT” Workshop, where the FAIR

principles were initially conceived, and the publication of the paper describing the FAIR principles [1]. A list with more detailed information on the workshops is available as supporting information.¹

All BYODs have played an important role in iteratively improving the structure of subsequent ones, as well as in facilitating the adoption and research on FAIR. As representative examples, the following subsections describe the inaugural BYOD – the Human Protein Atlas and MycoBase BYOD – and the series of BYOD workshops focused on linking rare disease registries. The BYODs evolved based on the trainers’ perception during the workshops and based on informal feedback from attendees. No direct data from any participant was collected.

The first BYOD workshop: Human Protein Atlas and MycoBase

The first data-focused BYOD workshop was held in Leiden, the Netherlands, on the 24th and 25th of June, 2014 [19, 20]. The BYOD, which focused on making data interoperable, brought together data owners from the Human Protein Atlas [21] and MycoBase [22] with Linked Data experts. It is important to note that the FAIR principles and, consequently, the concept of “FAIR” data were still under development by then. Therefore, this BYOD focused on creating “Linked Data”, which is a step towards having FAIR data.

The data owners needed to be familiar with their current internal data management structures, i.e. the database schema and data pipelines for creating and displaying entries. The Linked Data experts had a variety of backgrounds such as semantic web services [23] and integration platforms [24]. The main aim was to develop sample Linked Data to demonstrate the added value of interoperable data for facilitating answering research questions by reusing information from multiple resources.

The BYOD event started with a plenary training session providing an overview of Linked Data. Then, the attendees were split into two working groups, each of which aimed to develop a proof of concept centred around one of the data resources, driven by their own research questions. The Human Protein Atlas group focused on developing a subset of Linked Data from the Human Protein Atlas. This was then connected to WikiPathways data [25]. The MycoBase group linked their data with the content of ChEMBL through the Open PHACTS API [26, 27]. The Human Protein Atlas developers have used the experience of the event to develop their own RDF data release, heavily reusing the ontological model of neXtProt [28].

In this BYOD, it became clear there was a need to include preparation and follow-up meetings in the agenda of subsequent workshops. The experience gained by organisers provided insights for planning pre-BYOD training about the FAIR principles and for organising post-BYOD supporting sessions. Additionally, it highlighted the importance of publishing training materials that could be used at other BYOD events, promoting knowledge sharing and dissemination.

A series of annually recurring BYODs: rare diseases registries

Making rare disease resources interoperable and, thereby, preparing them for multi-source analysis is crucial since rare diseases occur at low frequency. In Europe, a disease is considered rare when it affects less than 5 in 10,000 individuals [29]. Ensuring the interoperability of rare diseases data is important because non-integrated data would likely be insufficient to support research or improvements in outpatient care. Therefore, each local data resource is of relatively limited value on its own, but may be highly valuable in combination with other data.

¹<https://doi.org/10.5281/zenodo.8155154>

Table 2. Overview of ‘Bring Your Own Data’ (BYOD) workshops organised from 2014 to September 2023.

Title	Date and Location	Focus
FAIR Principles Idealisation - Jointly Designing a Data FAIR-PORT	13 - 16 Jan 2014 - Leiden, the Netherlands	-
The first BYOD workshop	24 - 25 Jun 2014 - Leiden, the Netherlands	Data
The first Bring Your Own Data (BYOD) Workshop To Link Rare Disease Registries - First RD BYOD Workshop	26 - 27 Nov 2014 - Rome, Italy	Data
The first "green genetics" BYOD	21 - 22 Jan 2015 - Wageningen, the Netherlands	Data
The Bring Your Own Template (BYOT) workshop	20 Feb 2015 - Utrecht, the Netherlands	Data
The second RD BYOD Workshop	24 - 25 Sep 2015 - Rome, Italy	Management
FAIR Principles Paper Published	15 Mar 2016	-
The third RD BYOD Workshop	29 - 30 Sep 2016 - Rome, Italy	Data
The FAIR Data Hackathon	19 - 20 Oct 2016 - Utrecht, the Netherlands	Software
The Software Solution Provider BYOD	25 - 27 Oct 2016 - Leiden, the Netherlands	Software
The Bring Your Own Rett Syndrome Data workshop	1 - 3 Nov 2016 - Maastricht, the Netherlands	Data
How to Make Data FAIR for Open Science	15 - 19 May 2017 - Leiden, the Netherlands	Data and management
The plant phenotype BYOD and hackathon	30 May - 1 Jun 2017 - Ghent, Belgium	Software
The cancer genomics BYOD	6 - 8 Jun 2017 - Utrecht, the Netherlands	Data and software
The fourth RD BYOD workshop	21 - 22 Sep 2017 - Rome, Italy	Data
The DSM BYOD workshop	25 - 26 Sep 2017 - Delft, the Netherlands	Data
The fifth RD BYOD workshop	13 - 14 Sep 2018 - Rome, Italy	Data and management
The RIKILT/WUR BYOD workshop	22 Nov 2018 - Wageningen, the Netherlands	Data
BYOD FAIRification workshop at Leiden University Library	18 Jun - 2019 - Leiden, the Netherlands	Data
The sixth RD BYOD workshop	26 - 27 Sep 2019 - Rome, Italy	Data and management
The seventh RD BYOD workshop	1 - 2 Oct 2020 - Online	Management
The eighth RD BYOD workshop	30 Sep - 1 Oct 2021 - Online	Management
The ninth RD BYOD workshop	29 - 30 Sep 2022 - Online	Management
The World Duchenne Organization's FAIR Training Program	7 - 9 March 2023 - Online	Management
The tenth RD BYOD workshop	28 - 29 Sep 2023 - Rome, Italy	Management

The first BYOD for rare disease registries and biobanks² (RD-BYOD) was held in Rome, Italy, at Istituto Superiore di Sanità on the 26th and 27th of November, 2014. The RD-BYOD was attended mainly by RD-Connect partners [30], including rare disease data owners and software engineers with Linked Data expertise. The main focus was to train data owners in making rare disease patient registries and biobanks interoperable, while also identifying tools to be developed.

With support of the BYODs, the rare disease community quickly acknowledged the importance of data interoperability, and later the FAIR principles. In 2017, the International Rare Disease Research Consortium (IRDIRC) declared the FAIR guiding principles as a ‘recognised resource’ to “accelerate the pace of translating discoveries into clinical applications” [31]. Since 2019, the series of annually recurring RD-BYODs has been an intrinsic part of the annual summer school on Rare Disease Registries. Editions of the course have been approved by the International Conference On Rare Diseases and Orphan Drugs (ICORD) [32].

Over the years, the RD-BYODs evolved to alleviate the steep learning curve of FAIRification. For instance, trainers were instructed to avoid very technical terms that could confuse beginners or participants with different expertise. Additionally, the RD-BYOD has evolved in response to feedback from participants and advancements in FAIR procedures and technologies. For example, training on FAIRification project management for registry managers was added in 2016 and expanded in subsequent editions of the workshop. As a result, from 2017, priority was given to attendees who were involved in or actively planned to establish a rare disease registry, primarily within a European Reference Network (ERN) [33], shifting the focus from a data-focused to a management-focused structure.

The RD-BYOD has also been used to experiment with, get feedback on and disseminate the technical developments that support the rare diseases community. It also informs registry managers about the available tools and standards. Recent RD-BYODs have been adapted to reflect practical aspects of the rare diseases domain, such as including topics to address needs from patient organisations and ERNs. As an example, the EJP RD ontological model for “Common Data Elements”, its supporting tool [14], and the EJP RD ontological metadata model for rare disease patient registries, biobanks and catalogues [15] were presented to participants in the latest editions, together with hands-on sessions for demonstration. The experience acquired by RD-BYOD trainers has been embedded in guidance resources, such as a guide for data stewards to make European rare disease patient registries FAIR [34].

Discussion

The BYODs have benefited attendees and trainers in many ways. For trainers, the workshops have created a collaborative environment where the FAIR community gains new insights from the attendees while helping them deal with their FAIR(ification)-related needs. For example, researchers on FAIR use the open and flexible BYOD environment to test FAIR-related tools with attendees. Similarly, feedback and questions raised during BYODs have supported research on FAIR and FAIRification methods. For instance, research on goal-based FAIRification planning methods [35], assessment of RDF data [36], large-scale implementation of FAIR principles [37] and quality of modelling [38] has benefited from experience from recent BYODs.

Furthermore, lessons learnt from success and shortcomings of BYODs provide guidance on future research paths. To illustrate, the pre- and post-BYOD activities

²For the sake of readability, the “BYOD for rare disease registries and biobanks” is referred to as “RD-BYOD” in this subsection.

have underscored the iterative nature of FAIRification, where the target resource is initially addressed and then expanded by subsequent FAIRification efforts. For example, in the first FAIRification iteration, a subset of data concepts within a given dataset may be addressed, with subsequent iterations expanding the scope to encompass a larger set of concepts. Other challenges, such as solving the communication gap due to the interdisciplinary nature of FAIR and the diverse expertise of BYOD attendees, highlight the need for further research on such topics. Additionally, the difficulty in reaching consensus during conceptual modelling tasks [39], which are crucial in FAIRification [37], is another obstacle frequently encountered in BYODs.

For attendees, the workshops have aided the participating community by fostering the convergence of standards and tools. In this way, BYODs have become a valuable resource for advancing FAIR data practises. The BYODs' structure has inspired various FAIR training activities and courses, some of which are already offered by universities, other research institutes, or consortia as part of their research data management programmes (e.g. [40–42]). The Metadata for Machines workshop (M4M) [43] and the Three-Point FAIRification Framework (3PFF) [44] are also examples of a training frameworks that were inspired by the BYODs. Similarly, other FAIRification workflows and frameworks have embedded knowledge acquired by researchers who participated in the BYODs (as trainers or attendees). Examples of these include the generic workflow for the Data FAIRification process [4], the *de novo* FAIRification process of a registry for vascular anomalies [5], the FAIR in action framework for guiding FAIRification [45] and the FAIR Hourglass model for FAIRification and FAIR orchestration [46].

It is also noticeable that BYODs reflect the maturing of the FAIR community. While early BYODs used prototype tools designed to handle specific FAIRification tasks, recent BYODs have introduced more comprehensive tools that can cover a greater part of FAIRification. For instance, while the first BYODs for rare diseases reused generic tools for converting small datasets to linked data, recent ones introduced software systems that can automatically make data FAIR upon collection (e.g. Castor, MOLGENIS). Moreover, recent BYODs were capable of presenting more complex real-world FAIRification cases that lead to new insights and facilitated the retrospective FAIRification of a patient-led registry (e.g. The Duchenne Data Platform [47]).

For future training activities, we recommend combining different types of BYODs to tackle various tasks needed at different stages of FAIRification projects. Practically, a FAIRification project starts by creating a homogeneous basic knowledge about FAIR among all people involved, thus making the commitment and investment efforts clear to the whole FAIRification team. This can be done with management-focused BYODs. After the FAIRification project has been set up and its objectives have been identified, it is necessary to gather sufficient technical expertise, which can be supported by organising data- and software-focused BYODs.

For future BYODs, we plan to explicitly align the contents with the knowledge units mentioned in Appendix E of the FAIRsFAIR Teaching and Training Handbook [48]. Furthermore, we aspire to make our teaching materials themselves FAIR, in order to contribute to overcoming the shortage of FAIR expertise and to continue to keep learning as instructors.

Conclusion

Initially, BYODs aimed at making data interoperable by using available Linked Data technologies. Since their inception, BYODs have evolved and provided a collaborative space to develop FAIRification tools and more robust technologies. Additionally, BYOD workshops have become an important means of exchanging views and knowledge on FAIRification, and on informing researchers and managers on the benefits of FAIR.

Experience has shown that FAIR implementations are an effective approach to enable multi-source analysis, and the BYODs are a valuable asset in promoting the adoption of the FAIR principles in various domains. We will, therefore, continue to organise BYODs to accelerate the adoption and promotion of the FAIR principles.

Supporting information

S1 Table. Detailed table of the ‘Bring Your Own Data’ (BYOD) workshops organised from 2014 to September 2023. An extended version of Table 2, with additional information (i.e., description, links to past BYOD announcements) is available at <https://doi.org/10.5281/zenodo.8155154>

Acknowledgments

We would like to express our gratitude to the following individuals, institutions, and funding sources for their contributions to the realisation of the Bring Your Own Data (BYOD) workshops: the Bring Your Own Data Workshop To Link Rare Disease Registries (RD-BYOD) initiative received support from the RD-Connect project (funded from the European Community’s Seventh Framework Program under grant agreement n° 305444 “RD-CONNECT”), ELIXIR and ELIXIR-EXCELERATE (Grant number EU H2020 #676559), the Istituto Superiore di Sanità (ISS), the Leiden University Medical Center (LUMC), the University Medical Center Groningen, and the Dutch Techcentre for Life Sciences (DTL) between 2014 and 2018. Since 2019, the RD-BYOD has been funded by the European Joint Programme Rare Diseases (EJP RD) and its partners (European Union Horizon 2020 Research and Innovation Programme under Grant Agreement n° 825575), and we are grateful for their continued support. We would also like to acknowledge the Trusted World of Corona (TWOC) project funded by Health-Holland. Finally, we would also like to extend our appreciation to all the attendees and trainers who actively participated in the various BYODs, particularly representatives from patient organisations. Their engagement, insights, and expertise have significantly enriched our work and shaped its outcomes.

References

1. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*. 2016;3:1–9.
2. Lusher S, Mons B. Jointly designing a data FAIRPORT; 2014. Lorentz Center at Snellius. Available from: <https://www.lorentzcenter.nl/jointly-designing-a-data-fairport.html>.
3. da Silva Santos LOB, Burger K, Kaliyaperumal R, Wilkinson MD. FAIR Data Point: A FAIR-oriented approach for metadata publication. *Data Intelligence*. 2022; p. 1–21.
4. Jacobsen A, Kaliyaperumal R, da Silva Santos LOB, Mons B, Schultes E, Roos M, et al. A generic workflow for the data FAIRification process. *Data Intelligence*. 2020;2(1-2):56–65.
5. Groenen KH, Jacobsen A, Kersloot MG, dos Santos Vieira B, van Enckevort E, Kaliyaperumal R, et al. The de novo FAIRification process of a registry for vascular anomalies. *Orphanet journal of rare diseases*. 2021;16(1):1–10.

6. Mark D Wilkinson and Dominique Batista. FAIR Evaluation Services; 2023. <https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/>.
7. FAIRSharing Team and Data Readiness Group. FAIRAssist; 2023. <https://fairassist.org/#!/>.
8. FAIR Data Team. FAIRifier; 2023. <https://github.com/FAIRDataTeam/FAIRifier>.
9. Kersloot MG, Jacobsen A, Groenen KH, dos Santos Vieira B, Kaliyaperumal R, Abu-Hanna A, et al. De-novo FAIRification via an Electronic Data Capture system by automated transformation of filled electronic Case Report Forms into machine-readable data. *Journal of Biomedical Informatics*. 2021;122:103897.
10. van der Velde KJ, Imhann F, Charbon B, Pang C, van Enckevort D, Slofstra M, et al. MOLGENIS research: advanced bioinformatics data software for non-bioinformaticians. *Bioinformatics*. 2019;35(6):1076–1078.
11. W3C. Data Catalog Vocabulary (DCAT) - Version 2. World Wide Web Consortium (W3C); 2019. Available from: <https://www.w3.org/TR/vocab-dcat-2/>.
12. (W3C) W3C. Resource Description Framework (RDF); Accessed 2023. <https://www.w3.org/RDF/>.
13. Creative Commons. Creative Commons Attribution-NonCommercial 4.0 International License; 2023. <https://creativecommons.org/licenses/by-nc/4.0/>.
14. Kaliyaperumal R, Wilkinson MD, Moreno PA, Benis N, Cornet R, dos Santos Vieira B, et al. Semantic modelling of common data elements for rare disease registries, and a prototype workflow for their deployment over registry data. *Journal of biomedical semantics*. 2022;13(1):1–16.
15. EJP-RD. Metadata for EJP rare disease patient registries, biobanks and catalogs; 2023. <https://github.com/ejp-rd-vp/resource-metadata-schema>.
16. EJP-RD Consortium. International Summer School on Rare Disease Registries and FAIRification of Data; 2023. <https://www.ejprarediseases.org/event/international-summer-school-on-rare-disease-registries-and-fairification->
17. dos Santos Vieira B, Bernabé CH, Zhang S, Abaza H, Benis N, Cámara A, et al. Towards FAIRification of sensitive and fragmented rare disease patient data: challenges and solutions in European reference network registries. *Orphanet journal of rare diseases*. 2022;17(1):436.
18. DTL. Successful Hackathon to Make MOLGENIS FAIR; 2016. <https://www.dtls.nl/2016/10/30/successful-hackathon-make-molgenis-fair/>.
19. Roos M, Gray AJ, Waagmeester A, Thompson M, Kaliyaperumal R, Van Der Horst E, et al. Bring Your Own Data Workshops: A Mechanism to Aid Data Owners to Comply with Linked Data Best Practices. In: SWAT4LS; 2014.
20. Gray, Alasdair J G . First FAIRPORT-ELIXIR BYOD workshop; 2014. [Online]. Available at: <https://alasdairgray.github.io/posts/2014-07-11-first-byod-workshop>.

21. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based human protein atlas. *Nature biotechnology*. 2010;28(12):1248–1250.
22. Robert V, Vu D, Amor ABH, van de Wiele N, Brouwer C, Jabas B, et al. MycoBank gearing up for new horizons. *IMA fungus*. 2013;4:371–379.
23. Wilkinson MD, McCarthy L, Vandervalk B, Withers D, Kawas E, Samadian S. SADI, SHARE, and the in silico scientific method. In: *BMC bioinformatics*. vol. 11. BioMed Central; 2010. p. 1–5.
24. Williams AJ, Harland L, Groth P, Pettifer S, Chichester C, Willighagen EL, et al. Open PHACTS: semantic interoperability for drug discovery. *Drug discovery today*. 2012;17(21-22):1188–1198.
25. Slenter DN, Kutmon M, Hanspers K, Riutta A, Windsor J, Nunes N, et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic acids research*. 2018;46(D1):D661–D667.
26. Gray AJ, Groth P, Loizou A, Askjaer S, Brenninkmeijer C, Burger K, et al. Applying linked data approaches to pharmacology: Architectural decisions and implementation. *Semantic Web*. 2014;5(2):101–113.
27. Willighagen EL, Waagmeester A, Spjuth O, Ansell P, Williams AJ, Tkachenko V, et al. The ChEMBL database as linked open data. *Journal of cheminformatics*. 2013;5(1):1–12.
28. Lane L, Argoud-Puy G, Britan A, Cusin I, Duek PD, Evalet O, et al. neXtProt: a knowledge platform for human proteins. *Nucleic acids research*. 2012;40(D1):D76–D83.
29. Simone Baldovino M, Domenica Taruscio M, Dario Roccatello M. Rare diseases in Europe: from a wide to a local perspective. *Isr Med Assoc J*. 2016;18:359–63.
30. RD-Connect. RD-Connect; 2022. Retrieved May 3, 2023, from <https://rd-connect.eu/>.
31. IRDiRC. Current IRDiRC Recognized Resources; 2023. <https://irdirc.org/resources-2/irdirc-recognized-resources/current-irdirc-recognized-resources/>.
32. RD E, editor. International Summer School on Rare Disease Registries and FAIRification of Data. Rome, Italy: Istituto Superiore di Sanità; 2022.
33. European Commission. European Reference Networks; 2023. https://health.ec.europa.eu/european-reference-networks/overview_en.
34. van Damme P, Moreno PA, Bernabé CH, Ballesteros AC, Le Cornec CM, Vieira BDS, et al. A Resource for Guiding Data Stewards to Make European Rare Disease Patient Registries FAIR. Ubiquity Press. 2023;.
35. Bernabé C, Sales TP, Schultes E, van Ulzen N, Jacobsen A, da Silva Santos LOB, et al. A goal-oriented method for FAIRification planning. *Research Square*. 2023;(Preprint).
36. Zhang S, Benis N, Cornet R. Automated approach for quality assessment of RDF resources. *BMC Medical Informatics and Decision Making*. 2023;23(1):1–16.

37. dos Santos Vieira B, Bernabé CH, Henriques I, Zhang S, Camara AB, García JAR, et al.. Critical steps towards large-scale implementation of the FAIR data principles; 2023. Available from: <https://doi.org/10.5281/zenodo.7867293>.
38. Bernabé C, Queralt-Rosinach N, Souza VES, da Silva Santos LOB, Mons B, Jacobsen A, et al. The use of Foundational Ontologies in Bioinformatics. Research Square. 2022;(Preprint).
39. Neuhaus F, Hastings J. Ontology development is consensus creation, not (merely) representation. Applied Ontology. 2022;(Preprint):1–19.
40. Hettne, Kristina M and Verhaar, Paul and Companjen, Bas and Kaliyaperumal, Rajaram and Jacobsen, Annika and Burger, Kees and Sesink, Leonie. Bring Your Own Data FAIRification workshop 18 June 2019; 2023. Retrieved from <https://osf.io/8ecg2>.
41. University of Twente. FAIR Principles and the FAIRification process course; 2021. <https://osiris.utwente.nl/student/OnderwijsCatalogusSelect.do?selectie=cursus&cursus=202100258&collegejaar=2021&taal=en>.
42. Aktau A, Gambardella A, Hettne K, Ulzen van N. Final outcomes of the taskforce FAIRification as as Service; 2023. Available from: <https://doi.org/10.5281/zenodo.7546767>.
43. GO FAIR. Making FAIR Metadata; 2023. <https://www.go-fair.org/today/making-fair-metadata/>.
44. How to GO FAIR;. <https://www.go-fair.org/how-to-go-fair/>.
45. Welter D, Juty N, Rocca-Serra P, Xu F, Henderson D, Gu W, et al. FAIR in action-a flexible framework to guide FAIRification. Scientific data. 2023;10(1):291.
46. Schultes E. The FAIR hourglass: A framework for FAIR implementation. Fair Connect. 2023;1(1):13–17.
47. van Lin N, Paliouras G, Vroom E, t Hoen PA, Roos M. How patient organizations can drive FAIR data efforts to facilitate research and health care: A report of the virtual second international meeting on Duchenne data sharing, march 3, 2021. Journal of Neuromuscular Diseases. 2021;8(6):1097–1108.
48. Engelhardt C, Biernacka K, Coffey A, Cornet R, Danciu A, Demchenko Y, et al. D7.4 How to be FAIR with your data. A teaching and training handbook for higher education institutions. Zenodo; 2022. Available from: <https://doi.org/10.5281/zenodo.6674301>.