EXCELERATE Deliverable 5.3

Project Title:	ELIXIR-EXCELERATE: Fast-track ELIXIR implementation and drive early user exploitation across the life sciences	
Project Acronym:	ELIXIR-EXCELERATE	
Grant agreement no.:	676559	
	H2020-INFRADEV-2014-2015/H2020-INFRADEV-1-2015-1	
Deliverable title:	Bring Your Own Data (BYOD)	
WP No.	5	
Lead Beneficiary:	6 - NBIC	
WP Title	The ELIXIR Interoperability Backbone	
Contractual delivery date:	31 August 2019	
Actual delivery date:	22 May 2019	
WP leader:	Carole Goble; Chris Evelo	4 - UNIMAN; 32 - UL
Partner(s) contributing to this deliverable:	DTL; UNIMAN; UL; LUMC; UOXF; UMAAS; NBIC	

Authors and Contributors:

Rob Hooft (DTL), Carole Goble (UNIMAN); Chris Evelo (UL), Marco Roos (LUMC), Susanna Sansone (UOXF), Friederike Ehrhart (UMAAS), Mateusz Kuzak (DTL), Claudio Carta (Instituto Superiore di Sanità), Helen Parkinson (EMBL-EBI), Mascha Jansen (NBIC), Barend Mons (LUMC)

Reviewers:

Helen Parkinson (EMBL-EBI)

1.Table of contents

Horizon 2020 grant n. 676559

2. Executive Summary

A series of workshops have been conducted that have resulted in more FAIR data, FAIRer data repositories, more knowledge about FAIR data and FAIR data processes, and software to support making data FAIR. Similar workshops are getting traction outside of the EXCELERATE project too.

3. Impact

In at least six of the Bring Your Own Data meetings (BYODs) focusing on data, and five focusing on (repository) software, a total of more than 100 practitioners have been taught practical steps to make data interoperable by FAIR data modeling using linked data and ontologies. This has resulted in FAIRer data, and in the development of tools supporting data FAIR-ification. An especially large impact has been made in the domain of rare disease research by organising annual BYOD workshops for rare disease data managers together with that community and WP8 of EXCELERATE. A scientific paper is in preparation that details the experience with, and results of, BYODs. Several other international projects are now using BYOD workshops too, and the concept has also already been brought into practice in collaborations with industry partners.

4. Project objectives

No.	Objective	Yes	No
1	Objective 1: Define agreements on identifiers and machine processable (meta)data descriptions with data providers (WP3, WP6 to 9), for (i) data repositories and (ii) knowledge bases, using established technical and domain standards, working with global initiatives to ensure broader interoperability		X
2	Objective 2: Consolidate existing Interoperability Services to support: (i) machine processable identity, data formats, experimental reporting guidelines, knowledge representations, and (ii) resource operational practices for transparent releases, versioning, provenance, updates.		X
3	Objective 3: Implement data interoperability between prioritized resources, in priority areas.	X	
4	Objective 4: Run a programme of Bring Your Own Data (BYOD) “bootcamps” and coordinate with WP11. Build capacity with WP10: Data Nodes Network.	X	

Horizon 2020 grant n. 676559

5. Delivery and schedule

The delivery is delayed: Yes • No☑

6. Adjustments made

Not applicable.

7. Background information

Work package number	5	Start date or starting event:	Month 1
Work package title	**The ELIXIR Interoperability Backbone**		
Lead	Barend Mons (up to 1/3/2017) (NL) Chris Evelo (since 1/3/2017) (NL) and Carole Goble (UK)		
Participant number and person months per participant 1 – EMBL 12.00; 2 – UOXF 10.40; 3 - TGAC 10.60; **4 - UNIMAN 19.00**; 6 - NBIC 23.00; LUMC 4.00; UMAAS 10.00; DLO 1.00; 8 - CRG 2.20; 10 - IRB 13.00; 12 - BSC 15.00; 13 - CSIC 2.80; 17 - INESC-ID 6.90; 24 - UiO 4.00; 25 - SIB 6.00; 26 - CNRS 10.00; 30 - CNR 4.26; **32 - UL 6.00**; 34 - UOCHB 16.19; 38 - DTU 10.00; 45 - UU 20.00 (KTH 2.00; SU 10.00); 46 - HWU 11.00			
WP5 - The ELIXIR Interoperability Backbone [Months: 1-48] UL, EMBL, UOXF, TGAC, UNIMAN, NBIC, CRG, IRB, BSC, CSIC, INESC-ID, UiO, SIB, CNRS, CNR, UOCHB, DTU, UU, HWU Optimal Interoperability is attained when data access and use can be completely automated: programming and interfaces conform to standards that specify consistent syntax and formats; and data are associated with metadata and terminology identifiers and codes that support computational aggregation and comparison of information that reside in separate resources. As an exemplar implementation of this principle, The Interoperability Backbone in WP5 is data and Use Case driven, implementing FAIR principles, working in partnership with the custodians of the datasets. We will support cross-resource questions, for instance bridging genomic and phenotypic data for variant identification, using machine processable (RDF/XML) representations of the metadata. We use this as a reference for building Node capacity in skills and knowledge for data interoperability and access to technical infrastructure that supports data interoperability. For example: Rare disease data described with the ORDO (Orphanet Rare Disease			

Horizon 2020 grant n. 676559

ontology) to integrate rare and common disease (WP8); Different plant phenotypic data described with a variety of specialist and overlapping terms (e.g. Plant Ontology, Crop Ontology, Plant Trait Ontology) with sample environment descriptors drawn from the Environment Ontology (EnvO) and the eXtensible Experiment Markup Language (XEML) (WP7); Marine metagenomics data described using EnvO, XEML, plus descriptions to screen environmental metagenomic sequence datasets (e.g. FOAM), and the Genomic Standards Consortium's reporting checklists and provenance aspects(such as environment from which samples originate).
For the data resource and service management of Named and Core Resources (WP3) we emphasize self-described datasets and explicitly described and published life-cycle metadata, using machine processable representations and common APIs for accessing it. The approach taken here will be an iterative one, working on tasks in parallel, to ensure that continuous but stepwise improvements for interoperability are yielded, and forged in practice against expressed need. We will not attempt comprehensive perfection: instead we aim for a principled "Just Enough and Just in Time" interoperability "on demand", while raising the bar on general data service quality along with WP1. No standard is committed to without an example implementation.

The objectives for WP5 are:

1. Define agreements on identifiers and machine processable (meta)data descriptions with data providers (WP3, WP6 to 9), for (i) data repositories and (ii) knowledge bases, using established technical and domain standards, working with global initiatives to ensure broader interoperability.
2. Consolidate existing Interoperability Services to support: (i) machine processable identity, data formats, experimental reporting guidelines, knowledge representations, and (ii) resource operational practices for transparent releases, versioning, provenance, updates.
3. Implement data interoperability between prioritized resources, in priority areas.
4. Run a programme of Bring Your Own Data (BYOD) "bootcamps" and coordinate with WP11. Build capacity with WP10: Data Nodes Network.

Work Package Leads: Barend Mons (up to 1/3/2017) (NL) Chris Evelo (since 1/3/2017) (NL) and Carole Goble (UK)

Task 5.1: FAIR Principle Interoperability Implementation Agreements (52PM)
Many of the necessary services already exist across ELIXIR but need alignment and support. This task is aimed at identifying critical services, defining the interfaces between them, and developing a long-term integration strategy. By doing so, this will survey national (Node) practices and needs. Work is based on existing ELIXIR infrastructure and community emergent conventions and the forging of partners in and outside Europe. The WP partners have established track records of metadata management for biology datasets or metadata mechanisms, notably using Linked Data (e.g. Semantic Web) technologies for publishing self-described data (e.g. RDF), resolving identifiers(e.g. URIs) and defining ontologies.
For datasets we will agree:

Horizon 2020 grant n. 676559

- Practices of data management and data publishing; Managed APIs and message formats, with agreed APIs for access to dataset descriptors. We tackle data repositories and biological knowledge bases differently, reflecting their different content and their content lifecycles.
- Common exchange formats: such as RDF and XML data schemes.
- Common reporting guidelines: submission, curation and validation tools using data templates (example: ISATAB and tools), focusing on interoperability of standards via common data element mappings.
- Common terminologies: general (e.g. VoID, PROV for provenance); data type/community specific (examples: ORDO, Plant Trait Ontology); cross-community common elements (example: EnvO environment descriptors).
- Common APIs: for common data types.
- Best practices for publishing data as Linked Data, leveraging the EMBL-EBI's RDF Platform, LinkedISA, and resources of other platforms (example: IMI Open PHACTS Discovery Platform), as a semantic interoperability platform in addition to the use of APIs.

For biological knowledge bases the task will concentrate on common conventions for:

- Descriptions using common terminology, standard data formats, and mappings between common data elements and standard ontologies.
- Descriptions of the dependencies, curation and computational processes used to generate the current record for the biological entity, where appropriate.
- Good practices for publishing data as Linked Data, leveraging the EMBL-EBI's RDF Platform and resources of other platforms (example: IMI Open PHACTS Discovery Platform), as a semantic interoperability platform in addition to the use of APIs.

Partners: all partners will be involved in sprints, focussing on specific dataset combinations to achieve WP6 to 9 questions.
Managers/Core: UK, NL, EMBL-EBI, SE

Subtask 5.1.1: Use Cases (WP6 to 9) (22PM)

Outcome: Interoperability, common APIs and descriptors workable in the field. First on WP7 (Genomic and Phenotypic Data for Crop and Forest Plants) and WP8 (Rare Disease) followed by WP6 (Marine metagenomics). Jointly identify data-driven interoperability examples, cross-dataset questions, bottlenecks and common descriptions. Survey national (Node) practices and needs.

Subtask 5.1.2 Core and Named Resources (WP3) (15PM)

Outcome: Common APIs and dataset descriptors workable in the field. Focus on interoperability at the dataset level.
Workshops to co-produce common APIs and dataset descriptors leveraging proposed standards (example: W3C HCLS
Dataset descriptor). Contribute to dataset metrics/quality criteria for data service life-cycle management in WP3. Survey national (Node) practices and needs.

Horizon 2020 grant n. 676559

Subtask 5.1.3 Global engagement: international organisations and multilateral forums (15PM)
Outcome: Establish a leading role for ELIXIR internationally on this aspect and compatibility with other international interoperability initiatives. Engagement with European initiatives (examples: IMIs, RIs, EUDAT EUON), global initiatives (examples: NIH BD2K, GA4GH, Force11), commodity standards (examples: W3C, DataCite, ORCID, VIVO, ORE) and community standards (example: RDA). Feedback between external forces and ELIXIR resources, maintaining ELIXIR visibility in key meetings with interoperability initiatives.
Task 5.2 FAIR Interoperability Implementation Services (55PM)
Integrating complex datasets requires services to handle identifiers for data and biological concepts (phenotypes, diseases), and tools that allow users to map data between different sources and help users find and apply standards. Many of these tools exist. This task brings these services together.
Partners: UK, NL, EMBL-EBI, SE, FR
Subtask 5.2.1: Identity Management, Mapping and Tracking services (14PM)
Outcome: Make explicit the scope and limitations of identifiers, the mappings between identifiers for entities (example: Ensembl gene identifiers can be mapped to Uniprot identifiers) and provide identifier services used by data resources in the field. Work includes: identity authorities for specific data types and concept categories; identity resolution, identity mapping, and entity resolution.
Subtask 5.2.2: Reporting Guidelines, Formats, Controlled Vocabulary Services (8.2PM)
Outcome: The best of breed services assembled, organized into a coherent tool suite, and used in practice in WP3 and WP6 to 9. Workshops of service providers and users to "bake-off" alternatives and integrate complementary services.
Subtask 5.2.3: Dataset publishing for API interoperability (10.8PM)
Outcome: Common and standardised practices to dataset lifecycle management and release management, including: Distributed revision control (example: GIT), dependency management; and well described, validated and maintained APIs registered in catalogues (example: BioCatalogue, BioSharing).
Subtask 5.2.4: Biological knowledgebase publishing for Linked Data interoperability (12PM)
Outcome: Data published as Interoperable Linked Data for some biological knowledge bases. We will develop services for the creation and management of mappings, as first class artefacts, between data entities to describe the curation and computational processes used to generate the current record for the biological entity (leveraging work at EBI/SIB, DTL, SciLifeLabs and the UK).
Subtask 5.2.5: Sustainability of Interoperability Implementation Services (12PM)
Outcome: A strategy for sustaining key services.
Task 5.3: "Bring Your Own Data" (BYOD) & Capacity Building Workshops (105PM)
Implement data interoperability between resources for WP6 to 9 and data publishing of resources for WP3. Provide practical support and guidance through a programme of Bring Your Own Data bootcamps. BYODs are a mix of tutorials and hands on, practical

Horizon 2020 grant n. 676559

"hackathons" with specific datasets: database custodians' work with experts in semantic web and linked data technologies to make their data available in a FAIR way using machine processable (meta) data and APIs. During the BYOD Task 5.1's specifications, standards and data templates are refined in practice and Task 5.2's tools and services are exercised.
Partners: NL, UK, ,FR, PT, SI, SE, CZ, IT, ES

Subtask 5.3.1: Manage and Run BYOD Workshops (65PM)

Outcome: Organise and manage workshops synergistically with WP3 and WP6 to 9 work plans, as well as aligned with WP11 training activities. Monitor workshops and feedback to Tasks 5.1 and 5.2. After the first 6 months of the project we anticipate a BYOD every 3 months.

Subtask 5.3.2: Create and manage BYOD training materials (20PM)

Outcome: Materials for BYOD, updating in step with Tasks 5.1 and 5.2. Deploy BYOD materials on the WP11 TeSS Portal. Develop BYOD materials for Data Carpentry training (WP11).

Subtask 5.3.3: Data Node Capacity Building (20PM)

Outcome: Build capacity in the Data Nodes (WP10, Task 10.2) using 5.3 as a reference example for interoperability practice and the BYOD methodology rolled out across the Nodes. Have capability to independently run BYOD's for new or national datasets in one third of the Nodes.

8. Appendix 1: Bring Your Own Data

8.1 Introduction

Before ELIXIR EXCELERATE started, the new NL node in ELIXIR had developed a new concept for a workshop aiming to make data more interoperable using Linked Data. These workshops became known as BYOD: "Bring Your Own Data" workshops. In the EXCELERATE proposal, the new concept was embraced as one of the cornerstones of the Interoperability work package, which has during the start of the implementation of EXCELERATE resulted in the formation of the Interoperability platform in ELIXIR.

During the course of EXCELERATE, ELIXIR-associated researchers, who also have been teaching in the BYOD workshops, have been instrumental in defining needs for future research data, resulting in the FAIR[1] principles. Since the definition of the FAIR principles BYOD workshops have broadened in perspective to cover all four letters of the FAIR acronym --making data more Findable, Accessible, Interoperable and Reusable-- rather than focusing exclusively on the Interoperability.

This deliverable reports on BYOD workshops executed with the help of EXCELERATE and in order to realize the goals of EXCELERATE. It summarizes the impact of these workshops for users of ELIXIR, and also briefly describes the experience gained.

[1] The FAIR Guiding Principles for scientific data management and stewardship, M.D. Wilkinson *et al*. Scientific Data volume 3, Article number: 160018 (2016) doi: 10.1038/sdata.2016.18

This document attempts to be a brief summary of results of BYOD meetings, focused on the accomplishments of ELIXIR EXCELERATE. For more in-depth and generic description of BYOD methods and results, we refer to Jacobsen *et al.* [2019][2].

8.2 Methods

The main concept of a Bring Your Own Data meeting is to bring researchers who are producing a specific data set together with potential users of that data and with experts on data interoperability. In a 2-3 day workshop, the participants first work together on a generic interoperability model for the data set, first conceptually, then with existing ontologies (typically encoded by Linked Data technology). They then use that model to make their selection of sample data linkable and machine readable in terms of ontologies, and use this to answer pre-set scientific questions that require at least one other set of data. Where possible, this is followed by a generalization: using the data to answer questions that were not pre-set for the meeting, but which now have become tractable through the new interoperable data model. The general procedure is used in the rare disease domain and described in EXCELERATE Deliverable 8.2[3].

In order to make the actual workshop as effective as possible, teleconferences with the participants (one or two focussed on data owners and subject matter experts, one on data linking experts) are organised several weeks ahead of the meeting. In those teleconferences the aims for the workshop are set, and everyone is prepared for their tasks.

The time of the workshop itself is too short to perform a complete FAIR-ification of a larger data set: the focus is really on the right model for the data, and to test that model on an interesting subset of the data. The participants take this experience home where they can apply the new model to the entire data set and can use their FAIR data expertise in their other work. To support this process, a few weeks after the BYOD workshop another teleconference is held. This follow up teleconference is used to be able to give additional guidance from the data modeling experts to resolve issues that have come up when bringing the model into practice.

At each BYOD workshop there is also room for capacity building. There is space for participation of new data modeling experts, learning about the process and training them on the job to participate as full experts in future BYODs.

Besides the standardized process of web meeting-workshop-web meeting, several versions of Bring Your Own Data workshops have been developed during EXCELERATE. Most notable is the series of Rome workshops run in collaboration with the Rare Disease use case[4], that is now part of the annual summer school for rare disease data managers. Here, the preparatory webinars have been gradually replaced by plenary learning sessions during the summer school. Follow-up is also not limited to a teleconference. The workshops have been arranged as annual events where participants can return to continue their work. Furthermore, the community has started a GO FAIR implementation network to foster FAIR principles in the rare disease community. Their activities include assisting participants in organising a FAIR-ification project after the summer school.

[2] Annika Jacobsen et al. Manuscript in preparation. https://docs.google.com/document/d/1OS3sqyLFynmS7BlcAmLEhYMzSiWqu8oPpeuJjwws1nI/edit.
[3] https://zenodo.org/record/1452468#.XOaPfC2ZNTY
[4] https://www.go-fair.org/implementation-networks/overview/rare-diseases/

Horizon 2020 grant n. 676559

8.3 Results

The prime EXCELERATE BYODs are listed in the table below, together with some key indicators.

Table 1: Example BYOD events organized with the help of EXCELERATE. Note that during the period of EXCELERATE also other BYODs have been run, including supporting companies with their internal data.

Date and Location	Data sets or software handled	Number of experts / trainers	Total number of participants	Link to EXCELERATE Use Case
24-25/9, 2015, Rome, IT	RD Registries, BYOD for beginners	8-10	20-26	WP8
29-30/9 2016; Rome, IT	Rare Disease Registries	8-10	20-26	WP8
25-27/10 2016; Leiden, NL	Catalog Software[5]	3-4	8-10	
1-3/11 2016; Maastricht, NL	Rett syndrome data[6]	6	12	WP8
30/5-1/6 2017; Utrecht, NL	BrAPI Software[7]		22	WP7
6-8/6 2017; Utrecht, NL	Cancer Genomics data[8]	6	23	
16/2 2017; Utrecht, NL	Transmart software[9]			
18-22/9 2017; Rome, IT	Rare Disease Registries	8-10	20-26	WP8
13-14/9 2018; Rome IT	Rare Disease Registries	8-10	20-26	WP8
4/2, 11/2 2019; Utrecht, NL	SURF / DANS repositories	4	10	

[5] News brief: https://www.dtls.nl/2016/10/30/successful-hackathon-make-molgenis-fair/
[6] Full report: https://www.dtls.nl/2016/11/10/bring-rett-syndrome-data-workshop-summary/
[7] Full report: https://www.dtls.nl/2017/08/28/plant-phenotype-byodhackathon-ghent-report/
[8] Full report: https://www.dtls.nl/wp-content/uploads/2017/09/BYOD-OncoXL-June-2017-report.pdf
[9] Full report http://blog.thehyve.nl/blog/first-step-towards-fair-transmart

Horizon 2020 grant n. 676559

The outcomes of BYODs can be split up into a few different classes. The first is more FAIR data in FAIRer local repositories/databases. Secondly, BYODs have increased awareness of FAIR principles and what is involved in their implementation, both in terms of technologies and managerial aspects. We aimed for capacity building of new data aspects. Transfer of knowledge between FAIR experts and domain experts is an important aspect of BYODs. Thirdly, through the coordinated interaction with stakeholders, BYODs have had substantial influence on the development of FAIR-ification procedure and software supporting FAIR data management. As a further outcome can be seen the experience gained with the organisation of these kinds of events.

8.3.1 FAIR data, FAIR data models and FAIR databases

The primary motivation to organise each BYOD has been the production of new FAIR data sets or databases and their supporting database/registry software, in order to be able to do better science. BYODs provide a useful first step in making data FAIR. In some cases, such as when stakeholders are bioinformaticians (i.e. trained in IT), the BYOD is used directly to prepare data for a data integration experiment. In other cases, such as for rare disease data managers, a BYOD is a first step towards planning a FAIR-ification project. For different BYODs between one and a handful of data sources have been identified and addressed. During the BYOD these resources have been analysed, proper FAIR models have been created in collaboration between the data and subject matter experts on one side and the linked data and data modeling experts on the other side. Samples of the data that were selected for the meeting have been subjected to the newly developed models, thereby testing the models.

As part of several BYODs, also some example scientific questions have been answered. An example is investigation of genetic variation data, available in the RDF data model, of the rare neurological disease Rett syndrome at the BYOD in Maastricht. The answers to these questions were not considered part of the outcome of the BYODs: primary goal of these exercises was to test the data models and to show the data owners the value of the FAIR-ification process.

Through the education of researchers who are maintaining and curating important data sets, BYODs contribute to the ideal in which data is FAIRified as close to its source as possible, where the necessary metadata is most easily acquired. Software/repository BYODs make a contribution to a FAIR data world by facilitating the accumulation of more FAIR data: data in repositories supported by FAIR software adhere more to interoperable data standards at a global level.

In BYODs that handled frequently used data types, the models that have been produced are re-usable for other projects using the same data type. Currently most of these FAIR-ification results are not yet stored in a FAIR repository. However, FAIRsharing[10] already contains a list of community-developed data and metadata models.This means that at least data stewards can reuse and build upon this existing work.

[10] https://fairsharing.org/

8.3.2 FAIR data expertise

An important outcome of any BYOD is the increased level of understanding of (semantically) interoperable data models for researchers and data maintainers who can concretely make a difference in the FAIRness of these data sets.

The scalability of the process of BYODs is still a challenge. Very many existing data sets still require attention and using the current processes it will be impossible for the currently available experts in the field to support all the transformation. Each BYOD of course can build on the expertise and models from all previous BYODs, but true speedup will only be possible once more capacity is systematically built in every event.

Contrary to initial expectations it has proven to be hard to use exclusively the BYOD events to train new data experts to run their own BYODs: there has been interest from new people to participate and learn, but it has not been clear what the necessary prior knowledge was and also insufficient attention has been paid during the workshops to this aspect. Further efforts for training outside of the events themselves have started to support future scale. The experiences and expertise from the past BYODs have been captured in the document "Essential Steps of the FAIR-ification Process"[11]. In the beginning of the year 2018, an Implementation Network (IN) "FAIR Data Stewardship Curriculum"[12] has been formed under the GO TRAIN pilar of the GO FAIR initiative. The goal for the IN is to (i) create an open, Carpentries style curriculum[13] based on the "Essential Steps" document and (ii) to setup an infrastructure (e.g. worksop page template) necessary to teach workshops based on this curriculum.

We will be using that curriculum in combination with "on-the-spot" training in future BYOD workshops to build more expertise capacity.

8.3.3 Software supporting FAIR data and BYODs

After organising the first few BYODs it became clear that there are issues that come up in many FAIR-ification efforts. For example: many data sets have a tabular structure, and standard tasks in the FAIR-ification of such data are to make column and row headers unambiguous, define domain (data type and value) of the values in each column, and to make the relationships between columns in the data explicit. In order to make BYODs more efficient, a tool to aid in these FAIR-ification steps was developed based on OpenRefine[14] and a pre-existing RDF plugin. More generally, it was attempted to automate as much of the laborious work of FAIR-ification in BYODs so that the interactions could concentrate on the specifics of the data sets, particularly the interdisciplinary effort of eliciting the meaning of data as needed for large scale automated analysis.

The resulting tools are available as documented open source[15]. They are aimed for (i) reference software specifications for engineers, (ii) use in interdisciplinary teams involved in FAIR-ification, either in a BYOD setting or in more extensive FAIR-ification projects. They are currently not targeted for use outside such expert teams and cannot easily be used without expert guidance. They are still under development, using the feedback from

[11] Essential steps of the FAIRification Process; https://osf.io/avrys/ ; also findable through TeSS
[12] FAIR DS Curriculum IN; https://osf.io/qmsv3/
[13] The Carpentries Handbook | Lesson Development; https://docs.carpentries.org/topic_folders/lesson_development/index.html
[14] http://openrefine.org/
[15] https://github.com/FAIRDataTeam/

Horizon 2020 grant n. 676559

current and future BYODs and FAIR-ification projects for their refinement. Selected tools may mature further into independently usable FAIR-ification tools, although the use of their specifications in existing software has priority.

8.3.4 Event expertise

Most of these BYODs have been organised by or in collaboration with the ELIXIR-NL office. The biggest challenge encountered has been the funding for linked data experts that are not part of ELIXIR. Their participation could not be paid from the EXCELERATE grant. Participation of external experts has therefore been paid by different project funding, and during some BYODs experts have voluntarily donated their time.

The organisation of each BYOD has been a combination of special demands placed by the subject and the participants and a standard check list for the basic structure of the meetings.

8.4 Conclusions

Bringing people together in BYODs has proven to be a fruitful method to expose a significant group of people to detailed data modeling knowledge that can help make data FAIRer. This has made a welcome addition to the basic exposure to what has become known as the FAIR principles, which has been rising in the scientific community during the EXCELERATE project.

BYODs have proven to need support both from human experts as well as software tools. EXCELERATE has seen the first development of such tools which can speed up the repetitive curation tasks in a BYOD, further development of these and other tools will have the possibility to make future BYODs even more efficient. Every BYOD, however, also has new data modeling challenges, and for this reason it will always be necessary to have highly trained data modelers present at the workshops. This brings a challenge for scalability that was foreseen in EXCELERATE, however the training aspects of BYODs that had been planned have not worked as well as we would have wanted; possibly this was caused by the fact that the organisation needed to focus on the other aspects of the meeting. Nevertheless, the success of the Rome BYOD in the rare disease community, where it has become an annual event as part of a larger summer school, demonstrates its training value.

FAIR is a relatively new concept. The BYODs in EXCELERATE have been executed by the “interoperability” platform[16] of the ELIXIR organisation, and have also focused mainly on the interoperability of data. Some BYODs have worked on Findability and Accessibility aspects, but more of these aspects have been part of other workshops in ELIXIR, for example based on further developing schema.org[17].

8.5 Follow-up

BYODs were one of the first implementations of the idea of making data FAIR at the source, even before the FAIR principles had been formulated. BYODs will still be playing a role in the ELIXIR Interoperability platform in the second scientific programme

[16] https://elixir-europe.org/platforms/interoperability
[17] https://bioschemas.org

(2019-2023) of ELIXIR, where a task with the name “FAIR at the source” has been prioritized by the platform. Experience gained in EXCELERATE BYODs is also used in a 2018 ELIXIR implementation study (FAIRCDR[18]) with as goal to evaluate and improve the FAIRness of the ELIXIR Core Data Resources[19].

From the experience in ELIXIR EXCELERATE there are also several other initiatives and projects that are using BYOD meetings and planning to use BYOD meetings.

- In EXCELERATE, the Interoperability work package worked together with the rare disease use case on running BYOD meetings annually in Rome. These will be continued for the coming years in the European Joint Programme for Rare Diseases, EJP-RD[20]. The experience of ELIXIR, also a partner of the EJP-RD, will play a major role in those meetings. Rare disease patient organisations have also expressed interest in organising BYODs.
- The FAIRplus[21] project will use the experience gained in EXCELERATE BYOD meetings and use it in meetings with IMI partners.
- ELIXIR-NL have collaborated with life science researchers from other fields than the ELIXIR EXCELERATE use cases in workshops to make their data FAIR. This work will continue. Also, applications outside of the life sciences are showing interest to learn from our experience.
- Several small and large industry partners of ELIXIR-NL have shown interest in BYODs; at the receiving end of the workshop (making their own data interoperable) as well as organizers and source of expertise. A few workshops with commercial partners have taken place already during ELIXIR EXCELERATE, and since interest is still strong, this will only be extended in the future.

The association of ELIXIR with the Carpentries foundation[22] will offer possibilities for the necessary capacity building of BYOD expertise, thereby overcoming the lack of progress that has been made with this aspect thus far.

[18] https://elixir-europe.org/platforms/data/fairness-core-resources
[19] Durinx C, McEntyre J, Appel R et al. Identifying ELIXIR Core Data Resources [version 2; peer review: 2 approved]. F1000Research 2017, 5(ELIXIR):2422 doi: 10.12688/f1000research.9656.2
[20] http://www.ejprarediseases.org/
[21] https://fairplus-project.eu/
[22] https://elixir-europe.org/platforms/training/software-carpentry