EXCELERATE Deliverable 8.4

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:	Report on the ELIXIR workshop organized with the rare-disease communities	
WP No.	8	
Lead Beneficiary:	7 : CNIO	
WP Title	Use Case C: ELIXIR infrastructure for Rare Disease research	
Contractual delivery date:	31 October 2017 (Amendment 2)	
Actual delivery date:	5 November 2017	
WP leader:	Ivo Gut, Marco Roos	CRG, NBIC
Partner(s) contributing to this deliverable:	CNAG- ES, NTNU-NO, NBIC-NL, CNIO-ES, UL-SI, IT-CNR, DTU-DK, CNRS.-FR	

Authors and contributors: Brane Leskosek, Leslie Matalonga, Andrej Kastrin, Jure Dimec, Femke Francissen, Sergi Beltran, Eivind Coward, Hedi Peterson, Ivo Gut, Marco Roos, Susheel Varma

Table of content

1. Executive Summary 3

2. Project objectives 3

3. Delivery and schedule 3

4. Adjustments made 3

5. Background information 3

6. Appendix 1: Report on the ELIXIR workshops organized with the Rare Diseases communities 7

1. Workshops for and with the rare disease community 8

1.1 Semantic Web Applications and Tools for Healthcare and Life Sciences, SWAT4LS Conference 2016 - 5-8 December 2016, Amsterdam, NL 9

1.2 3rd International Rare Diseases Research Consortium Conference - 8-9 February, 2017, Paris, FR 10

1.3 ELIXIR All Hands Meeting - 21-23 March 2017, Rome, IT 10

1.4 Bioinformatics & Systems Biology 2017 conference - 4-5 April 2017, Lunteren, NL 11

1.5 RD-ACTION Workshop Co-hosted by DG SANTE: Using standards and embedding good practices to promote interoperable data sharing in ERNs - 26-27 April 2017, Brussels, BE 11

1.6 RD-Connect - NeurOmics - EURenOmics annual meeting - 1-3 May 2017, Berlin, DE 11

1.7 E-Rare Data Sharing and Harmonization Workshop - 3-4 May 2017, Berlin, DE 12

1.8 European Society of Human Genetics Conference 2017 - 27-30 May 2017, Copenhagen, DK 12

1.9 ELIXIR Innovation and SME Forum: Genomics, bioinformatics and health - Public-private partnerships in open data - 6-7 June 2017, Barcelona, ES. 12

1.10 5th International Summer School Rare Disease & Orphan Drug Registries - 18-20 September 2017, Rome, IT 13

1.11 Bring Your Own Data, BYOD, to Link Rare Disease Registries, 4th RD-Connect BYOD Workshop - 21-22 September 2017, Rome, IT 14

1.12 16th CRG Symposium: Seventh International Workshop on Genomic Epidemiology- 20-22 September 2017, Barcelona, ES 14

2. Future events 15

3. RD training needs survey 15

4. Results 16

5. Conclusion 19

Horizon 2020 grant n. 676559

1. Executive Summary

In this second report on training workshops we first describe the workshops for the past year and planned future workshops for the rare diseases (RD) domain within the context of the ELIXIR-EXCELERATE project. Secondly, the results of the ELIXIR Training - Rare Diseases Training capacity and needs survey 2017 - are presented with suggestions for future training focus for the RD domain within ELIXIR community. For the both parts the aims were (1) to identify key training contacts/subjects in the RD community, (2) to identify specific training needs in the RD community and (3) to spread specific knowledge needed in the RD community.

Events were either organised specifically with the Rare Diseases (RD) community, or a separate stream/session for the RD community was co-organised around other events. Training workshops and courses were delivered in partnership with WP11 'EXCELERATE Training Programme'.

2. Project objectives

With this deliverable, the project has reached or the deliverable has contributed to the following objectives:

No.	Objective	Yes	No
1	**Deliver world-leading data services for academia and industry :** Demonstrate, in partnership with the Rare Disease community, how aligned ELIXIR resources enable research, avoid fragmentation and support the development of sustainability models for resources created by the community research projects	X	

3. Delivery and schedule

The delivery is delayed: ☑ Yes • No

4. Adjustments made

Minor delays as result of the quality assurance process.

5. Background information

Background information on this WP as originally indicated in the description of action (DoA) is included here for reference.

Work package number	WP8	**Lead beneficiary**	8 - CRG
Work package title	Use Case C: ELIXIR infrastructure for Rare Disease research		

Horizon 2020 grant n. 676559

Start Month	1	End Month	48
Work Package Lead	Ivo Gut (ES) and Marco Roos (NL)		

Objectives

WP8 aims to empower actors involved in the development of new rare diseases therapies through the execution of the following specific objectives:
• Build the ELIXIR registry of data resources and analysis tools critical for the development of the rare disease research. (Task 8.1)
Continuous monitoring of resources and tools in Rare-diseases.
Implementation of a system for the generation of datasets adequate for the assessment of methods in the area of rare-diseases.
Implementation of the ELIXIR rare-disease portfolio in the ELIXIR registry.
• Implementation of a technical framework for the comparison and standardization of services useful for the rare-disease
communities. (Task 8.2)
• Collaboration with the rare-disease communities for the organization of training courses, workshops and jamborees.
(Task 8.3)

Description of work and role of partners

WP8 - Use Case C: ELIXIR infrastructure for Rare Disease research [Months: 1-48]

CRG, UNIMAN, NBIC, CNIO, IRB, BSC, NTNU, CNRS, CNR, UL, DTU

The International Rare Diseases Research Consortium (http://www.irdirc.org) established the ambitious goal ofdeveloping 200 new therapies by 2020. ELIXIR as a whole and in particular this Work Package is aligned with this effort. The overall objective of this Work Package (WP) is to address the needs of the rare diseases community through the instantiation of the ELIXIR resources described in WP1-5. These resources do not constitute a replacement of the current research projects organized around the rare diseases area. Indeed the aim is to empower them and to help in the sustainability of the resources created by these projects in the long term. This WP is organised around the actors that play a major role on the development of these new therapies. These actors

are the main users of the ELIXIR infrastructure: data generators and curators (usually personnel working in hospitals, genomics-based companies, and members of large research consortia), researchers (bioinformaticians, geneticists, and clinical doctors), diagnosis companies, CROs (usually SMEs), and the pharmaceutical industry among others

Task 8.1: The ELIXIR portfolio of data resources developed in collaboration with the rare diseases communities. (69.4PM)

Subtask 8.1.1 Monitoring of resources and tools. (25.4PM)

There is a wide range of data resources and analysis methods used in the rare-disease area. Many of those resources are provided by ELIXIR Nodes, for example the European Genome-Phenome archive (EGA) currently stores data from major research initiatives in rare diseases like the RD-connect project. In this subtask we will review the current data resources and

Horizon 2020 grant n. 676559

evaluate their usability and potential impact on the rare disease community. An important aspect of the evaluation will be the security of the data that is a key aspect in rare disease domain given the low frequency of the associated genomic variants in the population. One critical aspect of the development of the registry is to engage the different communities in the submission and rating of the tools. In this task we will work together with representatives of the major projects in the field of rarediseases to create a customized portfolio of ELIXIR tools and services devoted to assist them in the development of these new therapies. As an example we will ask for proposals of tools that serve to interpret the effect of genomics variants on a group of patients that belong to the same family. We strongly believe that this link between the end- users and the tools developers will help ELIXIR to understand better the problems that are actually facing the main actors in the rare diseases research and hence to better solutions.

The final outcome of this task will be the ELIXIR data resources and analysis tools useful to the rare disease

communities.

Partners: NO, ES, SI, IT, NL

Subtask 8.1.2: Creation of reference datasets adequate for the specific assessment of methods and standards in the area of rare-diseases. (30PM)
While the creation of these tools should stay as a priority for researchers, large scale projects, SMEs and the industry increasingly need access to benchmarked methods on which to build their analysis strategies. The evaluation of the methods requires the adequate selection of the datasets and benchmarking strategies. The systems for the selection of the datasets for the benchmarking have to be fast and effective to enable the continuous evaluation of the methods, as described in WP2. We will collaborate with the ELIXIR benchmarking strategy (WP2) to build the appropriate strategies for the selection of the datasets (subtask 8.1.1 above) and with the rare- disease communities
to implement the adequate quality reporting standards. Moreover we will integrate these pipelines in the ELIXIR benchmarking framework (WP2) to continuously monitor the selected methods with the newly generated datasets.
Partners: ES, DK, IT,FR, SI, UK

Subtask 8.1.3 Implementation of the ELIXIR rare-disease portfolio in the ELIXIR registry. (14PM)
The ELIXIR registry will be a reference for the research community (WP1), as it will reflect the quality and the realtime status of the services included on it. This registry will act as a one-stop shop for services provided by ELIXIR. The goal is to allow users from the different countries, communities and projects to discover which are the tools available at a given time, with the associated information about the community based rating (see WP2), instructions for correct use and associated examples We will encourage tools developers to adopt the EDAM standard to describe their tools and to share several metrics about the performance and usage of these of the tools (see description in WP1) Those services promoted as relevant by the end-users will be listed in a special section in the ELIXIR registry.
Partners: DK, ES, FR.

Task 8.2: Standardisation of rare disease services in collaboration with the RD communities. (36PM)
The ecosystem of RD services will inevitably be a combination of distributed and centralized resources, because of the sheer number of rare diseases and rare disease organisations, as well

Horizon 2020 grant n. 676559

as legal and ethical constraints between countries and communities. At the same time, because of the low frequency in the population, combining data across patient registries, biobanks, and -omics databases is the single most important way of getting new insights towards new treatments.
One of the most recurrent issues when attempting to perform research across resources is the lack of standards or the poor adoption of existing standards by RD stakeholders. Rare disease standards concern different types of data including genomic and phenotypic characteristics, causative genetic variation status, quality criteria, analysis protocols, supporting evidence and follow-up indicators. These problems will be analysed in workshops including experts in semantic web,
linked data technologies and rare-disease experts (see previous experiences and proposal in "Bring Your Own Data (BYOD) bootcamps", in WP5). The initial experience with this methodology (see 61) is that a critical bottleneck is the identification of the most appropriate terms and identifiers to annotate data for cross- resource questions. Based on this experience we aim to address two major 'white spots' in the available infrastructure for Rare-diseases: (i) the current infrastructure of the rare disease platform: RD-Connect, does not contain backbone services for functional interlinking,
(ii) a majority of RD sources are not equipped to provide data, metadata, and data updates using appropriate standard procedures. To address these needs we will work together with WP5, the rare-disease communities and the RD-Connect project to (i) deploy and test the services and guidelines for standardization 'at the source', (ii) provide standardized interfaces that Rare-disease communities can work with from a central location, (iii) build capacity in the RD community
by enabling them to work with these services themselves.
Partners: FR, ES, DK, NL.

Task 8.3: Training workshops targeting different user communities. (32PM)
In this task training workshops and courses will be delivered, in partnership with WP11 "EXCELERATE Training Programme". The training will be approached from two sides. First, in collaboration with the Train the Researcher task in WP11 we will train rare diseases' researchers in the use of relevant tools, standards and infrastructure produced by ELIXIR. Second, we will run "feedback workshops" in which those who are developing the methods will be exposed
directly to problems faced by the rare disease community. These userthons will help to shape the ELIXIR portfolio. The direct collaboration with WP11 Train the Researcher will ensure that researchers are trained to a high standard in state-of-the-art analysis techniques for rare disease data and that innovative training approaches developed in this task are applied elsewhere in ELIXIR.
Partners: UK, SI, NL.
Participation

Partner number, short name and effort: 4 - UNIMAN 6.00; 6 - NBIC 0.00 LUMC 6.00; 7 - CNIO 10.00; 8 - CRG 38.40; 10 - IRB 12.00; 12 - BSC 10.00; 22 - NTNU 12.00; 26 - CNRS 12.00; 30 - CNR 3.00; 32 - UL 15.00; 38 - DTU 12.00

Horizon 2020 grant n. 676559

6. Appendix 1: Report on the ELIXIR workshops organized with the Rare Diseases communities

Summary

In this second report on training workshops we first describe the workshops for the past year and planned future workshops for the rare diseases (RD) domain within the context of the ELIXIR-EXCELERATE project. Secondly, the results of the ELIXIR Training - Rare Diseases Training capacity and needs survey 2017 - are presented with suggestions for future training focus for the RD domain within ELIXIR community. For the both parts the aims were (1) to identify key training contacts/subjects in the RD community, (2) to identify specific training needs in the RD community and (3) to spread specific knowledge needed in the RD community. An important mission of EXCELERATE WP8 'ELIXIR infrastructure for Rare Diseases research', defined at the kick-off meeting on November 24, 2015 in Leiden, is to help the rare diseases community in different nodes to raise its infrastructure activities to the level of best practices in ELIXIR infrastructure, i.e. where rare diseases data, metadata, tools, and catalogues are shared, interoperable, and sustainable at the source, are conformant with a common set of principles and standards across participating countries/nodes and in agreement with the ELIXIR platforms. To achieve optimal knowledge exchange between the domains, workshops in this WP are typically co-organised with members from the rare diseases community and co-sponsored by relevant projects such as RD-Connect[1].

During this second reporting period a total of 12 different international events involving the participation of at least seven ELIXIR nodes (ES, NL, FR, IT, BE, DE and SI) and reaching more than 2000 experts and researchers in the RD field across Europe were organised. These events, coordinated by or in collaboration with the ELIXIR rare disease use case and the ELIXIR training platform, have also an impact in the the dissemination of all the possibilities ELIXIR infrastructure offers to the RD community through the different ELIXIR platforms.

Events were either organised specifically with the Rare Diseases (RD) community, or a separate stream/session for the RD community was co-organised around other events. Training workshops and courses were delivered in partnership with WP11 'EXCELERATE Training Programme'.

The delivery of workshops was complemented by the ELIXIR implementation study/RD-Connect proof-of-concept to test interoperability backbone components for enabling queries across rare diseases biobanks and registries.

Training capacity and needs for RD community were identified from the "ELIXIR Training: Rare Diseases Training capacity and needs 2017" survey augmented with "ELIXIR Training offerings, need and capacity survey 2016" data. In combined data set 112 respondents from 15 countries including EMBL-EBI node participated. Survey results will be used to plan future training events at the RD community.

[1] http://rd-connect.eu/

1. Workshops for and with the rare disease community

EXCELERATE members contribute to the workshops by advising in rare disease workshops, mediating them and organising the contribution of data experts, such as in the case of Bring Your Own Data events - BYODs (contributions of Marco Roos and Mascha Jansen for ELIXIR-NL). Based on the experience from the workshops, they take initiatives towards collaboratively developing protocols and tooling (e.g. the cross-project rare disease data linkage plan co-led by EXCELERATE WP8 co-lead Marco Roos), define new training goals (led by Brane Leskosek and Celia van Gelder for WP8/11), and further developing the Rare Disease Connect platform (led by WP8 co-leads Ivo Gut and Sergi Beltran). The objectives of the contributions are: (i) to stimulate and foster rare disease data resources to become FAIR at the source (Findable, Accessible, Interoperable, Reusable for humans and computers), (ii) to stimulate the rare disease community to share (meta)data via the RD-Connect platform and European infrastructures such as EGA. In the Table 1 is presented the summary of the workshops and events organised within the RD domain in the observed period (Table 1).

TABLE 1: LIST OF WORKSHOPS AND EVENTS

Nr.	Place	Date	Event	Participants		Topic
				Country	Num.	
1	Amsterdam, NL	5-8 Dec 2016	SWAT4LS Conference 2016	NL, BE, FR, DE, IT, UK, LU, ES, NO, CH, USA and more	95	Semantic Web Applications and Tools for Healthcare and Life Sciences
2	Paris, FR	8-9 Feb 2017	3rd International Rare Diseases Research Consortium Conference		200	Boosting health care and life science research on rare diseases RD-Connect ID-Cards of biobanks and registries
3	Rome, IT	21-23 Mar 2017	ELIXIR All Hands Meeting	all ELIXIR members + AT	200	Poster WP8 Workshop: "Data Interoperability" Collaboration with RD-Connect
4	Lunteren, NL	4-5 Apr 2017	BioSB 2017 conference	NL, DE, BE, SI	410	ELIXIR and RD-Connect workshop.
5	Brussels, BE	26-27 Apr 2017	RD-ACTION Workshop Co-hosted by DG SANTE	BE, NL		Interoperability, FAIR data and ERNs
6	Berlin, DE	1-3 May 2017	RD-Connect,NeurOmics, EURenOmics annual meeting	DE, ES, NL	100	ELIXIR and RD-Connect
7	Berlin, DE	3-4 May 2017	E-Rare Data Sharing and Harmonization Workshop	NL		
8	Copenhagen, DK	27-30 May 2017	European Society of Human Genetics Conference 2017	NL, ES,	500 - 1000	"Data Interoperability", ELIXIR and RD-connect
9	Barcelona, ES	6-7 Jun 2017	ELIXIR Innovation and SME Forum	ES	100	ELIXIR, personalized medicine and industry
10	Rome, IT	18-20 Sep 2017	5th International Summer School Rare Disease & Orphan Drug Registries	IT, NL		Interoperability and FAIR data
11	Rome, IT	21-22 Sep 2017	BYOD, to Link Rare Disease Registries	IT, NL		BYOD workshop
12	Barcelona, ES	20-22 Sep 2017	16th CRG Symposium	ES	150	Genomic Epidemiology

Horizon 2020 grant n. 676559

1.1 Semantic Web Applications and Tools for Healthcare and Life Sciences, SWAT4LS Conference 2016 - 5-8 December 2016, Amsterdam, NL

Program

The Conference brought together researchers, developers, and users, from many fields to discuss goals, current limits and real experiences in the use of Semantic Web technologies in health care and the life sciences, particularly this edition was so structured: a tutorials day (5th), main conference (6th and 7th), and a Hackathon day (8th) the last day. Among other events there was:

- Tutorial: **“FAIR Data and Data Stewardship”**

  During the 9th Edition was done a “FAIR Data and Data Stewardship tutorial” organised in two main parts:

  - A first part with presentations with the aim to ensure that participants see the rationale for each of the FAIR Principles, and in particular, understand the added value-proposition of achieving the Principles in the context of research initiatives and data stewardship.
    - The second part was a guided, hands-on in which participants went through the process of making data FAIR
- Plenary presentation: **“Registries of domain-relevant semantic reference models help bootstrap interoperability in domains with fragmented data resources”**

  During the presentation the requirements and subsequent design decisions that we had chosen to pursue during a still ongoing plan to make rare disease registries linkable at the source were shown.

  The plan was guided by experiences gained from a number of Bring Your Own Data workshops (BYODs) in the rare disease domain.

- Flash presentation and Poster: **"The organisation of Bring Your Own Data (BYOD) workshops to make life science data linkable at the source"**

  The afermentioned poster was selected for a "flash" presentation.

  We presented the general roadmap of the BYOD workshop even though each BYOD is uniquely tailored to the needs of the organiser. BYODs contain a preparatory phase with at least two webinars for data owners and domain experts, an execution phase for the BYOD itself, and a follow-up phase to foster the results of the BYOD by teleconferences with participants. A BYOD is also a learning experience that helps domain experts to endorse the approach in their domain.

1.2 3rd International Rare Diseases Research Consortium Conference - 8-9 February, 2017, Paris, FR

Program

Plenary presentation: **"Boosting health care and life science research on rare diseases by creating a robust infrastructure of independently Findable, Accessible, Interoperable, and Reusable (FAIR) biobanks, registries, and molecular data resources"** presented a 'rare disease data linkage plan' endorsed by stakeholders in the rare disease community and infrastructure experts. Stakeholders are committed to make rare disease resources FAIR by humans and computers at the source. FAIR principles are, besides ELIXIR, supported by several other infrastructures, projects and initiatives such as RD-Connect, CORBEL, BBMRI, Dutch projects FAIR-dICT and ODEX4All, and rare disease patient organisations. The RD data linkage plan provides recommendations for data annotation and exchange, and tooling.

Plenary presentation: **European Perspective on Sharing – Omics Data for Personalized Medicine in Rare Diseases** presented the actual European infrastructures and stakeholders for rare diseases (ELIXIR, EURORDIS, Orphanet, RD-connect, etc.), the importance of data sharing in secure environments and the need of benchmarking strategies

Poster: **"RD-Connect ID-Cards of biobanks and registries: making RD data Findable, Accessible, Interoperable and Reusable"**

The RD-Connect ID-Cards aims at concentrating sparse information on RD patients in one unique source by showing the number of samples/cases included in biobanks and registries, providing a first gateway towards more intense data sharing, and increasing the integration of biobanks and registries.

Furthermore, the RD-Connect-supported data linkage plan was presented and discussed in a number of the ERN break-out sessions.

1.3 ELIXIR All Hands Meeting - 21-23 March 2017, Rome, IT

Program

The third ELIXIR All Hands meeting took place 21-23 March 2017, bringing together members of the ELIXIR community from across the ELIXIR Nodes, and collaborators from partner organisations, in order to review ELIXIR achievements and activities so far and discuss plans for the future. The All Hands also acts as the ELIXIR-EXCELERATE Annual General Meeting.

During this event different **activities were presented:**

- Posters:
  - "Boosting genotype-phenotype and translational research on rare diseases by establishing Findable, Accessible, Interoperable and Reusable data resources through data linking technologies and Bring Your Own Data Workshops".
  - "Rare diseases data resources and tools survey".
- Workshop: "Data Interoperability": ELIXIR-CHARME Workshop: FAIR data and data stewardship in ELIXIR: How to write your own FAIRy tale.

- Flash talk: WP8- remote real-time visualisation of human rare disease genomics data (RD-Connect) stored at EGA: In this flash talk, we explained how we will provide real-time access and rendering of individual donor alignment files stored at the EGA (EMBL-EBI ELIXIR node) by a genome browser, such as Genome Maps or the Integrated Genome Browser.

1.4 Bioinformatics & Systems Biology 2017 conference - 4-5 April 2017, Lunteren, NL

Program

- **ELIXIR workshop: "Data interoperability & Disease"**
  - At this workshop the rare disease data linkage plan was presented.
- Poster "**Boosting genotype-phenotype and translational research on rare diseases by establishing Findable, Accessible, Interoperable and Reusable data resources through data linking technologies**"

1.5 RD-ACTION Workshop Co-hosted by DG SANTE: Using standards and embedding good practices to promote interoperable data sharing in ERNs - 26-27 April 2017, Brussels, BE

Program and output

The workshop had the overall aim to analyse and demonstrate how ERNs and their associated experts can use standards to optimise the utility and reusability of clinical data, by sharing the state of the art of rare disease data standardisation practices

During session three we presented the state of the art in linking data, including the following topics (i) the FAIR Data concept, (ii) what is FAIR data, (iii) what does this mean in the ERN context, through practical, Interactive demonstration.

1.6 RD-Connect - NeurOmics - EURenOmics annual meeting - 1-3 May 2017, Berlin, DE

Program

RD-Connect is an integrated platform for rare disease research that includes integrated analysis tools for whole human genome and exome data (https://platform.rd-connect.eu). RD-connect is tightly connected with ELIXIR and the ELIXIR-EXCELERATE RD use case through, for example, the storage of files (EGA, EBI-EMBL ELIXIR node) and by accessing the catalogue of RD resources through bio.tools (ELIXIR tools platform).

During this meeting, one training workshop, four poster and a plenary presentation were presented.

Training workshop: RD-Connect platform: The RD-Connect genomics analysis platform is live and being used by researchers across the world for gene discovery and accelerating diagnosis. This training session provided an opportunity to learn about how the platform can be used for rare disease research and how researchers can contribute data.

Four abstracts were accepted and presented as poster:

- Bring Your Own Data workshop a joint effort to promote and support FAIR RD-Registries
- Boosting genotype-phenotype and translational research on rare diseases by establishing Findable, Accessible, Interoperable and Reusable data resources through data linking technologies
- Strategy to support and improve quality of rare diseases registries
- Quality of whole exome sequences in RD-Connect

Plenary presentation:Platform development and integrated tools

1.7 E-Rare Data Sharing and Harmonization Workshop - 3-4 May 2017, Berlin, DE

Program

- Poster: “Bring Your Own Data workshop a joint effort to promote and support FAIR RD-Registries”
- The session four of the workshop was organized as a “Hands on” experience on FAIRifier data
- Panel discussions included contributions by WP8 co-leads Sergi Beltran (RD-Connect platform) and Marco Roos (FAIR data approach).

1.8 European Society of Human Genetics Conference 2017 - 27-30 May 2017, Copenhagen, DK

Program

Three abstracts have been submitted and accepted, two within the poster section - Omics/Bioinformatics

- Poster: “Boosting genotype-phenotype and translational research on rare diseases by establishing Findable, Accessible, Interoperable and Reusable (FAIR) data resources through data linking technologies.”
- Poster: Semi-automated generation of custom clinical genomic reports for rare diseases
- E-Poster: “Bring Your Own Data workshop is an excellent tool to promote the establishment of Findable, Accessible, Interoperable, and Reusable rare disease registries”

1.9 ELIXIR Innovation and SME Forum: Genomics, bioinformatics and health - Public-private partnerships in open data - 6-7 June 2017, Barcelona, ES.

Program

The aim of this Innovation and SME forum was to showcase to companies the free data resources and services that are available through ELIXIR Spain and ELIXIR Europe more generally.

During the first day, the ELIXIR rare disease Use Case through its application in personalized medicine was presented.

1.10 5th International Summer School Rare Disease & Orphan Drug Registries - 18-20 September 2017, Rome, IT

Program

There were two events organized at the end of September 2017 with the intent of promoting the establishment of Findable, Accessible, Interoperable and Reusable (FAIR) rare disease registries, in compliance with the IRDiRC and EU Recommendations.

The Summer School was organized by the National Centre for Rare Diseases (ISS) in collaboration with the Office for External Relations and Centre for International Affairs (ISS), RD-Connect, ELIXIR, RD-Action, EURORDIS. EpiRare, European Reference Networks and ICORD.

The objectives of the course were to assist the attendees:

I. in describing the resources needed for the establishment of a rare disease registry;
II. in describing the features of successful strategies to ensure:
    A. long term sustainability of the registry,
    B. data quality
    C. Compliance with FAIR data principles

In order to achieve these objectives the course was organized as an interactive session where speakers had a time slot for plenary presentation and discussion and/or practical exercises involving participants.

Moreover, for the third consecutive year, there were interactive small-group exercises in accordance with the Problem-Based Learning (PBL) methodology. This year particularly the One Day Problem Based Learning (ODPBL) was used, in which the groups must solve a problem in one day; the structure of ODPBL is the same of PBL. PBL is a highly interactive and learner-centred approach where learning occurs by working in a small group assisted by a facilitator to develop a solution to a problem. Scientific articles, expert lectures, consultations and feedback were the learning resources used to support the students in finding a solution to the problem. One of the ODPBLs was dedicated to FAIR data stewardship, which pertained to the ELIXIR contribution.

At the Summer School there were more than thirty two participants selected on the base of participant's background, role with reference to registry activities, and involvement in ERNs (5th International Summer School programme: http://www.iss.it/cnmr/?lang=1&id=2739&tipo=3).

1.11 Bring Your Own Data, BYOD, to Link Rare Disease Registries, 4th RD-Connect BYOD Workshop - 21-22 September 2017, Rome, IT

Program

BYOD to Link Rare Disease Registries took place in Rome at the Istituto Superiore di Sanità and the event was co-organised with RD-Connect, ELIXIR, RD-Action, EURORDIS. EpiRare, European Reference Networks, ICORD

A the BYOD there were more than twenty eight participants selected, as well as for International Summer School, on the base of participant's background, role with reference to registry activities, and involvement in ERNs (5th International Summer School programme: http://www.iss.it/cnmr/?lang=1&id=2739&tipo=3).

Before the BYOD workshop, online documents and preparatory webinars were organized to introduce FAIR principles, Semantic Web and Linked Data principles.

This year the focus was on (i) FAIRification with partially prepared semantic models, (ii) FAIR data tools, and (iii) FAIR data management.

The workshop was a hands-on experience. The attendees, divided into small groups, worked with FAIR data experts. Each group had to make the sample data of one of the four selected datasets, FAIR under the guide of their trainers. A the end to show the potentialities of data linkage, a cross resource question has been done against the four sample dataset which had been converted into FAIR resources by the four groups with the help of their IT-Trainers.

Both events, the 5th edition for the Summer School and the 4th edition for the BYOD workshop, generated a lot of interest and positive feedback.

1.12 16th CRG Symposium: Seventh International Workshop on Genomic Epidemiology- 20-22 September 2017, Barcelona, ES

Program

The 16th CRG Symposium: Seventh International Workshop on Genomic Epidemiology took place at the PRBB Auditorium and gathered speakers to discuss about the progress in sequencing and with other technologies for biomolecular analysis, and their applications in research and clinical settings.

Solutions have been presented for the accumulation, handling and interpretation of huge data sets, including the identification of rare and common genetic variants associated with disease, functional evaluation of genetic variation, understanding of gene networks and epigenomic phenomena in health and disease, pharmacogenomics, gene-gene and gene-environment interactions.

One plenary session: Sergi Beltran, CNAG-CRG, ES The BBMRI-LPC call to sequence 900 Rare Disease exomes: a successful transnational collaborative initiative with EuroBioBank and RD-Connect

One abstract have been submitted and accepted:

- Poster: Semi-automated generation of custom clinical genomic reports for rare diseases

2. Future events

A number of future training events are planned that will further enable establishment and more optimal maintenance of rare disease registries through specific f2f and e-learning courses with the help of WP11 and widespread and scaled up with the help of WP10.

- **Semantic Web Applications and Tools for Healthcare and Life Sciences, SWAT4LS Conference 2017**
- **International Summer School on: "Rare Disease and Orphan Drug Registries"**The next International Summer school will be structured in two parts a first part of three days with PBL and a second part with the hands on, BYOD.
    - The course will be held on 10-14 September 2018.
- **ELIXIR-All Hands 2018 Meeting,** 4-7 June 2018, Berlin, Germany
- **ELIXIR Innovation and SME Forum on Data-Driven Innovation in Rare Diseases and Personalised Medicine** 14-15 November 2017, Paris, France
- **ELIXIR-EXCELERATE WP8 face-to-face meeting for the rare disease community**, date TBD. In this meeting the interaction between the rare disease community and ELIXIR is discussed.
- **BioSB course: Managing and Integrating Life Science Information (4th edition)**: Approaches using Linked Data and Semantics (coordinators: Katy Wolstencroft, Marco Roos, Celia van Gelder), January 2018
- **EuRRECa Workshop on Rare Diseases Registries** (ENDO-ERN), Istituto Superiore di Sanità, February 8-9, 2018
- **International Sample & Data Banking course**, Istituto Ortopedico Rizzoli, Spring 2018

Bilaterally we have reached an agreement between RD-Connect and ELIXIR Slovenia node that ELIXIR-SI eLearning Platform (EeLP) developed within WP11 will be used for pilot/selected RD-Connect training events for RD community. Based on the results we will expand our further collaboration.

3. RD training needs survey

The survey "ELIXIR Training: Rare Diseases Training capacity and needs 2017" was executed in the first half of 2017. The execution was delayed, because we wanted to include as many nodes as possible, and align the survey well with the previously executed general ELIXIR Training survey, executed by WP11. The reason of alignment was also due to the possibility of data comparison between both surveys and increase reliability of the collected data. This will allow us to make a reliable comparison of results between general training needs in life science information and RD community training needs. Based on the results of the survey a new course/workshop(s) will be prepared in the next reporting period and where possible aligned with events in the rare

disease domain. The courses/workshops will use e-learning and other training techniques successfully tested in the ELIXIR training platform within WP11 Train the Researcher (TtR), Train the Trainer (TtT) and eLearning subtasks. All course materials will also be available through the RD-Connect project. The results of the survey are presented in the next section below. The survey was applied using the SurveyMonkey tool and administered to selected contacts within RD Community and through ELIXIR TrC group.

4. Results

The questionnaire comprised 11 blocks of activities, namely: sequence bioinformatics, system biology, structural bioinformatics, proteins and proteomics, high throughput sequencing, statistics, computational science, resource management, data management, health informatics, and training and teaching. Each block contained various items to describe each particular activity (e.g., sequence bioinformatics include the following activities: biological databases, alignment algorithms, multiple alignments, protein families etc.).

We augment the “ELIXIR Training: Rare Diseases Training capacity and needs 2017” data set with “ELIXIR Training offerings, need and capacity survey 2016” to get a more clear picture of the results and to expand sample size. In the first (main) survey the number of respondents equals 36 (from 12 countries) while in the second survey there were 76 respondents (from 15 countries including EMBL-EBI node). The actual combined sample size was thus 112 participants.

The figures below represent (i) events ran in 2015 and 2016 (Figure 1), (ii) events planned in 2017 and 2018 (Figure 2), and (iii) events needed in the future (Figure 3). Data are represented as heatmaps where each particular row refers to a country and each column refers to a block of activities. The color of a cell represents frequency of aggregated activities according to color key.

Horizon 2020 grant n. 676559

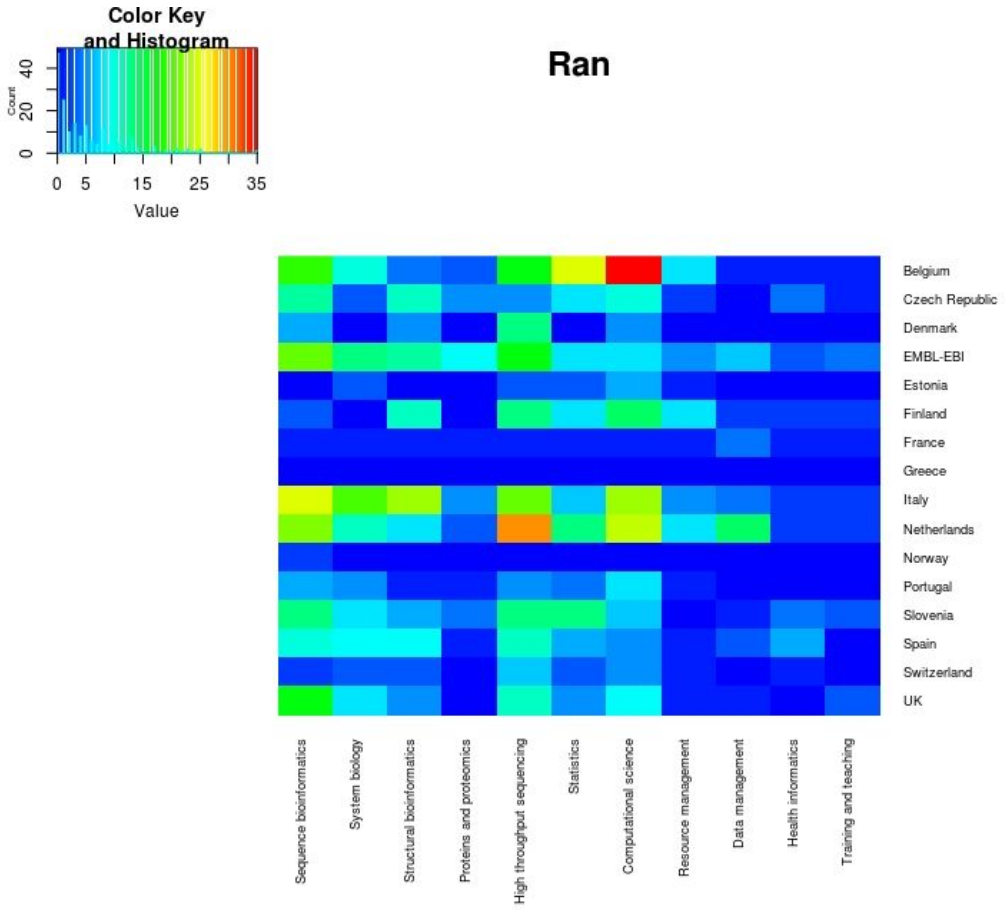

Figure 1: Heatmap of events ran in 2016

From Figure 1 we can observe that the most active countries are Netherlands, Italy, and Belgium, followed by EMBL-EBI and Slovenia. On the tail of the distribution are Greece and Norway. Most popular block of activities are high-throughput sequencing, computational science, and sequence bioinformatics while the least popular are health informatics and training and teaching

Horizon 2020 grant n. 676559

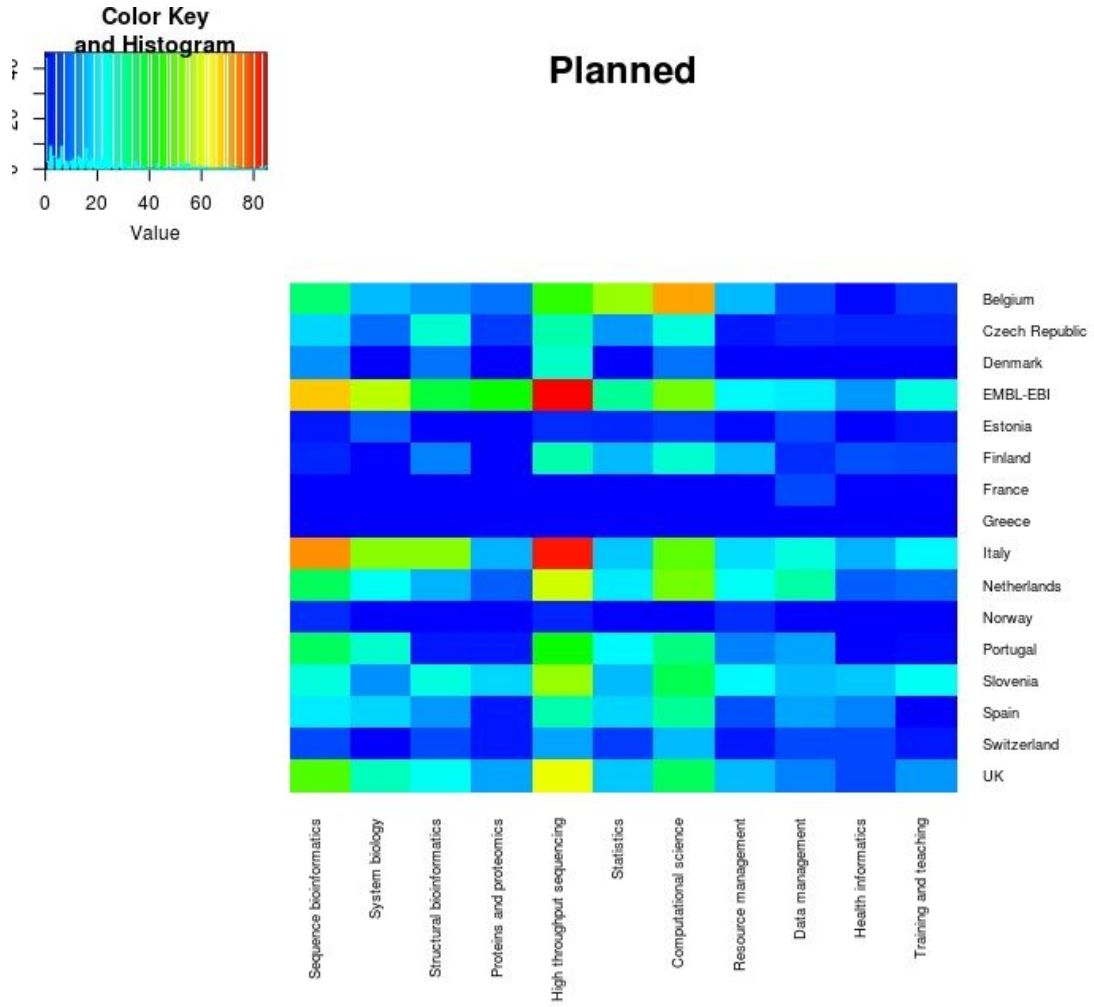

Figure 2: Heatmap of events planned for 2017 and 2018

According to planned activities, the most important respondents are EMBL-EBI, Italy and Netherlands, while Greece, France, and Norway exhibit low frequency of planned activities. Similar to ran activities, the most popular planned activities were high-throughput sequencing, computational science, and sequence bioinformatics, while health informatics and training and teaching are on the tail of the distribution.

Horizon 2020 grant n. 676559

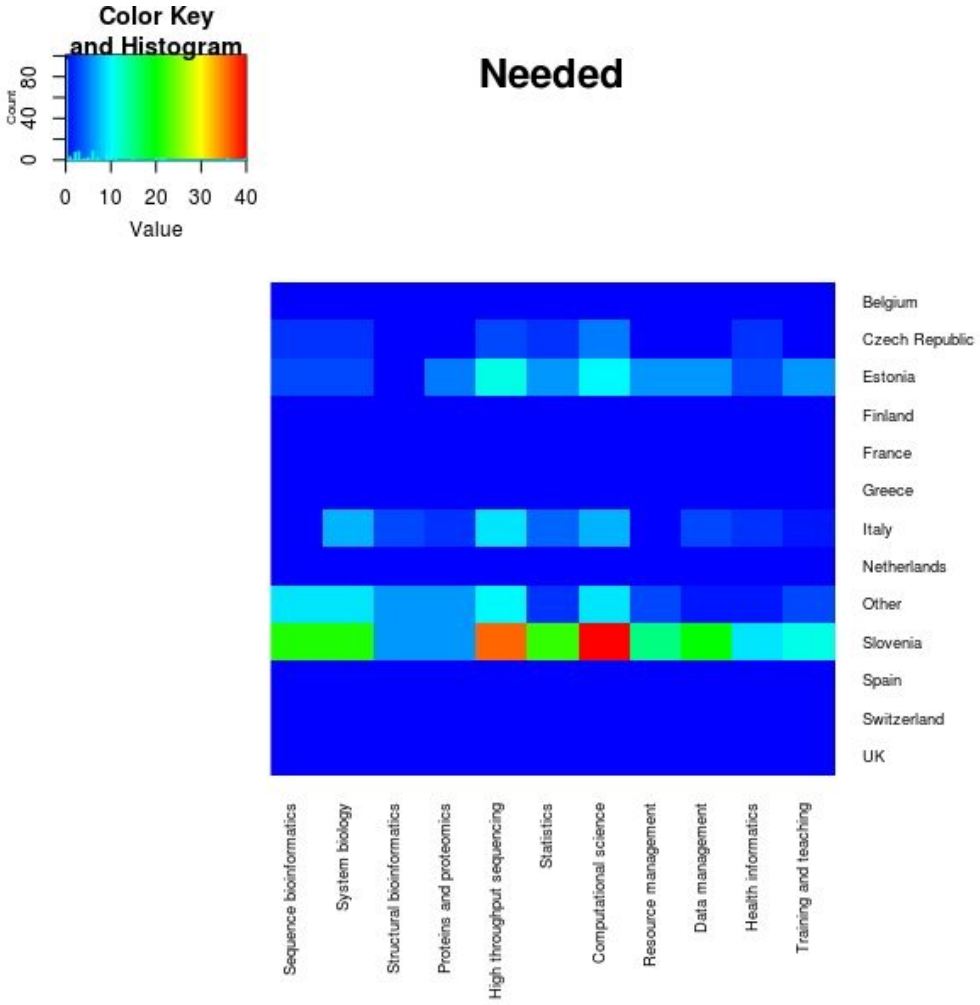

Figure 3: Heatmap of events needed in future

From Figure 3 above we can identify blocks of activities that are needed in the future. According to the results, there is an urgent need to provide more courses in computational science, high-throughput sequencing and system biology. Structural bioinformatics and health informatics deserve less attention. Countries that express greatest interest for new block of activities are Slovenia, Estonia and Italy. For instance, in Slovenia there is an urgent need for the training in “computational science” (especially in the fields of Python programming, machine learning, and big data analysis) and high throughput sequencing (in the fields of RNA-Seq and exome sequencing).

5. Conclusion

During the next reporting period the needs expressed via the survey will be evaluated against planned training events. The results will help to identify where new training events are appropriate, or where awareness of existing training events should be increased.

Horizon 2020 grant n. 676559