EXCELERATE Deliverable D8.7

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:	Options for sustainability of RD-Connect within the scope of ELIXIR	
WP No.	8	
Lead Beneficiary:	CNAG-CRG	
WP Title	Use Case C: ELIXIR infrastructure for Rare Disease research	
Contractual delivery date:	31 May 2019	
Actual delivery date:	23 May 2019	
WP leader:	Ivo Gut and Marco Roos	8 - CRG; 6 - LUMC (LTP)
Partner(s) contributing to this deliverable:	CNAG-CRG, NBIC (LUMC), EMBL (ELIXIR-HUB)	

Authors and Contributors:

Ivo Gut, Sergi Beltran, Marco Roos, Niklas Blomberg, Serena Scollen, Leslie Matalonga, Gary Saunders.

Reviewers:

None

1.Table of contents

Horizon 2020 grant n. 676559

2. Executive Summary

RD-Connect[1] is an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. Among other, RD-Connect has built three integrated online systems open to the RD Community: Genome-Phenome Analysis Platform[2] (GPAP), for analysis, interpretation and sharing of omics data, the Registry & Biobank Finder[3], a global directory of rare disease patient registries and biobank and Sample Catalogue[4], which helps researchers find rare biosamples stored in biobanks.

RD-Connect acknowledged early in its programme that it would be necessary to sustain these three main assets, among other, beyond the funding period of the project. For this reason, several initiatives were undertaken to ensure the sustainability of these resources. In many of those, BBMRI-ERIC[5] and ELIXIR[6] were invited to participate. The RD-Connect GPAP has been, and is, central to EXCELERATE WP8 developments, the ELIXIR Rare Disease Community and quite a few ELIXIR Implementation Studies. Since the RD-Connect project has been key for the EXCELERATE WP8, D8.7 has provided support to identify sustainability options for RD-Connect, within ELIXIR and beyond.

Most of the efforts have focused on the GPAP, and while basic operation funding are partially covered until 2023, EXCELERATE has helped to identify topics that need to be further developed to ensure long-term sustainability. Some of the topics identified include containerization of pipelines, running analysis on the cloud and federation of national/regional RD-Connect nodes while providing added value to data stored at the EGA. Altogether, this would provide a useful initial demonstrator for the MEGA+ and the 1M genomes declaration of the EU states[7].

3. Impact

RD-Connect GPAP will be sustained in the following 3-5 years by providing services to the RD Community such as the European Reference Networks that will use the platform as their main data reanalysis portal to collate around 19.000 existing exomes and genomes in the context of the Solve-RD project.

The RD-Connect GPAP has currently 694 users (clinicians / researchers) from 218 different groups and more than 15 European countries.

[1] https://rd-connect.eu/
[2] https://rd-connect.eu/what-we-do/omics/gpap/
[3] https://rd-connect.eu/what-we-do/phenotypic-data/rb-finder-for-registries/
[4] https://rd-connect.eu/what-we-do/biosamples-data/sample-catalogue/
[5] http://www.bbmri-eric.eu/
[6] https://elixir-europe.org/
[7] https://ec.europa.eu/digital-single-market/en/news/eu-countries-will-cooperate-linking-genomic-databases-across-borders

Horizon 2020 grant n. 676559

4. Project objectives

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

No.	Objective	Yes	No
1	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	

5. Delivery and schedule

The delivery is delayed: Yes • No☑

6. Adjustments made

Not applicable

7. Background information

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

Work package number	8	Start date or starting event:	month 1
Work package title	**Use Case C: ELIXIR infrastructure for Rare Disease research**		
Lead	Ivo Gut (ES) and Marco Roos (NL)		
Participant number and person months per participant 4 - UNIMAN 6.00; 6 - NBIC 0.00 LUMC 6.00; 8 - CRG 38.40; 9 - CIPF 2.66; 12 - BSC 10.00; 22 - NTNU 12.00; 26 - CNRS 12.00; 30 - CNR 6.39; 32 - UL 15.00; 38 - DTU 12.00; 47 - FPS 4.54			

Horizon 2020 grant n. 676559

The International Rare Diseases Research Consortium (http://www.irdirc.org) established the ambitious goal of developing 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.

Objectives

WP8 aims to empower actors involved in the development of new rare diseases therapies through the execution of the following specific objectives:

1. 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.
2. Implementation of a technical framework for the comparison and standardization of services useful for the rare-disease communities. (Task 8.2)
3. Collaboration with the rare-disease communities for the organization of training courses, workshops and jamborees. (Task 8.3)

Work Package Leads: Ivo Gut (ES) and Marco Roos (NL)

Description of work and role of partners

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

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 evaluate their usability and potential impact on the rare disease

Horizon 2020 grant n. 676559

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 rare- diseases 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 real-time 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.

Horizon 2020 grant n. 676559

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 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.

8. Appendix 1: Options for sustainability of RD-Connect within the scope of ELIXIR

8.1 Background

RD-Connect[8] is an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. The RD-Connect platform was built within the scope of a 6-year project funded by a FP7 grant from the European Commission. The funding for the project started in November 2012 and ended in October 2018.

Among other, RD-Connect has built three integrated online systems open to the RD Community:

- Genome-Phenome Analysis Platform[9] (GPAP) for analysis, interpretation and sharing of omics data to diagnose patients and discover new disease genes. The GPAP has been, and is, central to EXCELERATE WP8 developments, the ELIXIR Rare Disease Community and quite a few ELIXIR Implementation Studies.
- Registry & Biobank Finder[10], a global directory of rare disease patient registries and biobanks.
- Sample Catalogue[11], which helps researchers find rare biosamples stored in biobanks.

RD-Connect acknowledged early in its programme that it would be necessary to sustain these three main assets, among other, beyond the funding period of the project. For this reason, several initiatives were undertaken to ensure the sustainability of these resources, including face-to-face meetings, telephone calls and application to competitive funding. In many of those, BBMRI-ERIC[12] and ELIXIR were invited to participate.

To keep the RD-Connect members together and aligned with other stakeholders in Europe, the RD-Connect EMC agreed to launch the RD-Connect Community[13], a unique, international community bringing together experts from various disciplines and other stakeholders, including ELIXIR and BBMRI. The RD-Connect Community was officially established in July 2018 and will ensure a long-term collaboration within main rare disease stakeholders. However, solutions concerning short and long term sustainability of the resources developed within the RD-Connect project were needed.

8.2 Report

[8] https://rd-connect.eu/
[9] https://rd-connect.eu/what-we-do/omics/gpap/
[10] https://rd-connect.eu/what-we-do/phenotypic-data/rb-finder-for-registries/
[11] https://rd-connect.eu/what-we-do/biosamples-data/sample-catalogue/
[12] http://www.bbmri-eric.eu/
[13] https://rd-connect.eu/news-article/launch-of-the-rd-connect-community/

Horizon 2020 grant n. 676559

The meetings and calls in which ELIXIR and BBMRI-ERIC participated were used to identify that while the GPAP is more aligned with ELIXIR objectives and plans, the Sample Catalogue and the Registry & Biobank Finder are more aligned with BBMRI-ERIC. This report therefore refers solely to the sustainability plan of the RD-Connect GPAP. EXCELERATE has supported the development of such sustainability plan.

The RD-Connect GPAP has been key for the following EXCELERATE developments:

- Deployment of standards and validation datasets for variant calling pipelines (Laurie et al., 2016[14])
- Providing public reference genomic datasets for benchmarking rare disease activities within OpenEBench[15] (w/EXCELERATE WP2) (Milestone 8.2).
- Engagement of the RD-Community to build the ELIXIR registry of data resources and analysis tools critical for the development of the RD research[16] (Deliverable 8.1).
- Providing a specific use case and specifications for the bio.tools API to list and retrieve tools description used in the context of the RD-Connect GPAP[17] (w/EXCELERATE WP1).
- Establishment of a framework for quality assessment of genomic data (Deliverable 8.6).
- Containerising the RD-Connect variant calling pipeline following GA4GH standards [18] (w/ EXCELERATE WP4).

Furthermore, the RD-Connect GPAP has, or is, a key resource in the following ELIXIR Implementation Studies[19]:

- Rare Disease infrastructure implementation study (2019-2021)[20]
- Development of Architecture for Software Containers at ELIXIR and its use by EXCELERATE use-case communities (2019)[21]
- Implementation study for the integration of ELIXIR-IIB in ELIXIR Rare Diseases activities (2017-2018)[22]
- Remote real-time visualization of human rare disease genomics data (RD-Connect) stored at the EGA (2017-2018)[23]
- Integrating ELIXIR Luxembourg into ELIXIR activities (2017-2018)[24]

[14] 10.1002/humu.23114
[15] https://openebench.bsc.es/submission/repository/repositoryList.php
[16] https://rare-diseases.bio.tools
[17] https://RD-Connect.bio.tools
[18] https://github.com/inab/Wetlab2Variations
[19] Implementation studies are short technical projects, funded by the Member State contributed ELIXIR core budget, with the aim to link national services and prepare for long term sustainable Infrastructure Services (https://elixir-europe.org/about-us/implementation-studies)
[20] https://drive.google.com/file/d/10o4EMWpVsQoKRCUEKo-VhbcLkGiMe05z/view
[21] https://docs.google.com/document/d/1ScReM5En2CQrKU4VPWUllJTt0rCV4qgA7-HqKJKpoS0/edit?usp=sharing
[22] https://elixir-europe.org/about-us/implementation-studies/integrating-elixir-italy-elixir-activities
[23] https://elixir-europe.org/about-us/implementation-studies/visualization-aligned-genomics-data
[24] https://elixir-europe.org/about-us/implementation-studies/integrating-elixir-luxembourg

The RD-Connect GPAP will continue providing services beyond the original RD-Connect funding period thanks to projects that are using it as their main data collation and analysis resource or that will build on it to provide further functionalities and integration. CNAG-CRG[25] and other RD-Connect partners are part of Solve-RD[26], a Horizon 2020 project that kicked-off in January 2018 which is collating around 19,000 existing exomes and genomes using the RD-Connect GPAP as the main reanalysis portal used by the European Reference Networks. Solve-RD will also generate hundreds of additional datasets through new genome sequencing and RNA sequencing and will add new functionalities to the system and integration with other resources in the project.

On the infrastructure side, the >100M Euro H2020 European Joint Program on Rare Diseases[27] (EJP-RD, 2019-2023) has included the RD-Connect GPAP as one of its key infrastructures. CNAG-CRG, and other RD-Connect partners, are beneficiaries in the EJP-RD proposal, which also includes some funding for the Sample Catalogue and the Registry and Biobank Finder. Furthermore, ELIXIR (with its third linked parties) is also part of the EJP-RD and will bring in several resources. Funding will enable further development and integration between relevant RD resources. Collaboration between RD-Connect, ELIXIR and BBMRI-ERIC is envisioned to continue in the future to ensure further RD-Connect usage and sustainability.

The RD-Connect GPAP was selected an ELIXIR-ES resource in and is receiving funding equivalent to 24PM per year from the ISCIII (through Instituto Nacional de Bioinformática, ELIXIR-ES) since 2018. This funding is secured until 2020. On another note, ELIXIR has two ongoing Implementation Studies in which RD-Connect is involved to deploy its dockerized analysis pipeline[28], connect with EGA's GA4GH htsget server to visualize aligned genomic data and to integrate with tranSMART[29].

As part of the sustainability plan, the team has also increased the awareness and visibility of the RD-Connect GPAP. IRDiRC has included the RD-Connect GPAP as a recognised resource[30] and the E-Rare JTC2018[31], EJP-RD JTC2019[32] and the German BMBF RD calls[33] recommended or encouraged the usage of the RD-Connect GPAP. Discussions are underway with ELIXIR to apply to become an ELIXIR Core Data Resource or an ELIXIR Deposition Database for Biomolecular Data.

Further to the main RD-Connect GPAP instance, two regional instances have been deployed with the same code. One for the Undiagnosed Rare Disease Programme of Catalonia[34] and the other for the Navarra 1000 Genomes Project[35]. These nodes and the

[25] https://www.cnag.crg.eu/
[26] http://solve-rd.eu/
[27] http://www.ejprarediseases.org/
[28] https://docs.google.com/document/d/1ScReM5En2CQrKU4VPWUllJTt0rCV4qgA7-HqKJKpoS0/edit?usp=sharing
[29] https://drive.google.com/file/d/10o4EMWpVsQoKRCUEKo-VhbcLkGiMe05z/view
[30] http://www.irdirc.org/research/irdirc-recognized-resources/current-irdirc-recognized-resources/
[31] http://www.erare.eu/joint-call/1st-joint-call-european-joint-programme-rare-diseases-jtc-2018
[32] http://www.erare.eu/joint-call/1st-joint-call-european-joint-programme-rare-diseases-jtc-2019
[33] http://www.eucerd.eu/?page_id=542
[34] http://www.urdcat-cat

main RD-Connect GPAP instance could be used as an example towards a distributed, federated system of RD-Connect nodes, therefore supporting sustainability of the whole network. Basic discovery or full data sharing could be enabled based on the user profile/permissions. These developments could also serve the MEGA project and the declaration of the 1M genomes[36].

The work done for this deliverable also identified other areas in which ELIXIR could provide support for sustainability. While these still need to be properly assessed, some of the identified topics are:

- provide computing resources to the GPAP to enable federated genomic analysis
- provide support in deploying, and running, the standard GPAP variant calling and annotation pipeline in ELIXIR nodes
- provide support in federating RD-Connect systems horizontally and vertically with the EGA for data storage and access through the ELIXIR AAI (building on outcomes from WP9 and WP5)

The roadmap for the next 3-5 years will be to deploy RD-Connect local instances and work toward federated systems to ensure long term sustainability and serve the MEGA project and the declaration of the 1M genomes[37].

8.3 Conclusion

Since the RD-Connect project has been key for the EXCELERATE WP8, D8.7 has provided support to identify sustainability options for RD-Connect, within ELIXIR and beyond. Most of the efforts have focused on the GPAP, and while basic operation funding are partially covered until 2023, EXCELERATE has helped to identify topics that need to be further developed to ensure long-term sustainability. Some of the topics identified include containerization of pipelines, running analysis on the cloud and federation of national/regional RD-Connect nodes while providing added value to data stored at the EGA. Altogether, this would provide a useful initial demonstrator for the MEGA+ and the 1M genomes declaration of the EU states.

[35] http://nagen1000navarra.es
[36] https://ec.europa.eu/digital-single-market/en/news/eu-countries-will-cooperate-linking-genomic-databases-across-borders
[37] https://ec.europa.eu/digital-single-market/en/news/eu-countries-will-cooperate-linking-genomic-databases-across-borders