Consolidation of the ELIXIR RD community as a key player in the international field

Leads: Sergi Beltran (ELIXIR-ES) and Marco Roos (ELIXIR-NL)

Application made in response to the Request for Proposals published 3rd August 2018.

Abstract

The international RD community is invested in working collaboratively across borders in order to bring diagnoses and therapies to patients. Achieving this goal require access to an international RD infrastructure bundling data repositories, resources and tools. Building upon previous implementation studies and the ELIXIR-EXCELERATE project this proposal will contribute to the development of the international RD-oriented data analysis infrastructure with special focus on aligning with other international initiatives and integrating interoperable RD resources. This proposal will strengthen collaborations between the RD-Community, ELIXIR nodes and the ELIXIR platforms benefiting their services and resources in terms of FAIR data principles. This implementation study will impact on RD research by making available those resources to the international RD-Community and by training RD researchers in their skills to use analysis tools for the better diagnosis and later treatment of RD patients.

Index

Partners

Node	Name of PI	Institution	Role (Lead or Member)	Person Months
ES1	Sergi Beltran	CNAG-CRG	Co-Lead (WP1), member (WP2)	1.5 (WP1)
ES1	Ivo Gut	CNAG-CRG	Member (WP1)	In kind
ES2	Joaquin Dopazo	FPS	Co-Lead (WP2), member (WP1)	0.25 (WP1) + 2 (WP2)
ES3	Laura Furlong	IMIM	Member (WP2)	0.5 (WP2)
ES4	Salvador Capella-Gutierrez	BSC	Member (WP2)	0.5 (WP2)
ES5	Fátima Al-Shahrour	CNIO	Member (WP2)	In kind
			ES	**4.75**
FR1	Marc Hanauer	Orphanet	Co-Lead (WP2), member (WP1)	0.25 (WP1) + 2 (WP2)
FR1	Charlotte Rodwell	Orphanet	Member (WP2)	In kind
FR1	Ana Rath	Orphanet	Member (WP2)	In kind
FR2	Christophe Béroud	AMU	Member (WP2)	In kind
FR2	David Salgado	AMU	Member (WP2)	In kind
			FR	**2.25**
HU1	Balint L. Balint	UD	Co-lead (WP4), member (WP1), member (WP2)	0.25 (WP1) +1.5 (WP4)
HU2	Endre Barta	UD	Member (WP2), member (WP4)	In kind
			HU	**1.75**
IT1	Matteo Chiara	UNIMI	Member (WP2)	0.5 (WP2)
IT2	Federico Zambelli	UNIMI	Member (WP2)	In kind
IT3	Rita Casadio	UNIBO	Member (WP2)	0.5 (WP2)
IT4	Pier Luigi Martelli	UNIBO	Member (WP2)	In kind
IT5	Castrense Savojardo	UNIBO	Member (WP2)	In kind
IT6	Giacomo Tartari	UNIBO	Member (WP2)	In kind
IT7	Luana Licata	UNIROMA	Member (WP2)	In kind
IT8	Anna Marabotti	UNISA	Member (WP2)	In kind

IT9	Claudio Carta	ISS	Member (WP3)	In kind
IT10	Allegra Via	UNIROMA	Member (WP3), member WP4	0.5 (WP4)
			IT	**1,5**
LU1	Dietlind Gerloff	UL	Co-lead (WP3), member (WP1)	1.5 (WP1) + 1 (WP3)
LU2	Venkata Satagopam	UL	Member (WP3)	1.5 (WP3)
LU3	Regina Becker	UL	Member (WP4)	In kind
			LU	**4**
NL1	Marco Roos	LUMC	Co-lead (WP1 and WP3)	1 (WP1)
NL2	Chris Evelo	NBIC	Member (WP2), member (WP3)	0.5 (WP2) + 1.5 (WP3)
NL3	Martina Summer-Kutmon	UM	Member (WP2)	In kind
			NL	**3**
SE1	Jessica Lindvall	NBIS	Member (WP4)	0.5 (WP4)
			SE	**0.5**
SI1	Brane Leskošek	uni-lj	Co-lead (WP4), member (WP1)	0.25 (WP1) +1.5 (WP4)
			SI	**1.75**
UK1	Michael Sternberg	CISBIO	Member (WP2)	0.5 (WP2)
UK2	Christine Orengo	UCL	Member (WP2)	In kind
			UK	**0.5**
CZ	Hana Pergl Šustková	IOCB	Member (WP3)	In kind
DE1	Björn Grüning	Freiburg University	Member (WP4)	In kind
EBI1	Maria Martin	EMBL-EBI	Member (WP2)	In kind
EBI2	Henning Hermjakob	EMBL-EBI	Member (WP2)	In kind
NO1	Finn Drabløs	NTNU	Member (WP2, WP3, WP4)	In kind
NO2	Eivind Hovig	UIO	Member (WP2, WP3, WP4)	In kind
NO3	Inge Jonassen	UIB	Member (WP2, WP3, WP4)	In kind
PT1	Pedro Fernandes	IGC	Member (WP4)	In kind
PT2	Daniel Faria	IGC	Member (WP3)	In kind

	Total	**20**
Work period	Starting from May 2019 for a period of 2 years (M24)	

Background

The international rare disease (RD) community is invested in working collaboratively across borders in order to bring diagnoses and therapies to patients, in alignment with IRDIRC[1] vision of enabling all people living with a rare disease to receive an accurate diagnosis, care, and available therapy within one year of coming to medical attention. Achieving this goal will require a higher level of organisation efficiency as well as fundamental changes to the way science is conducted, shared, and applied towards the care of RD patients. Therefore, in alignment with the RD-Community roadmap, we will build on previous work performed in the context of ELIXIR-EXCELERATE WP8[2] and different ELIXIR implementation studies[3] to contribute to the international RD infrastructure efforts for RD research. We will ensure the coordination and linkage with existing infrastructures, projects and initiatives (WP1), contribute on building a global bioinformatics infrastructure for research (WP2), improve interoperability of data resources in terms of FAIR principles (WP3) and provide training for the RD community (WP4).

This proposal will work towards interoperability and integration of ELIXIR Core Data resources[4] and services supported by ELIXIR nodes and useful for RD research EGA (ELIXIR-ES), ArrayExpress (ELIXIR-EBI), UniProt (ELIXIR-CH), ChEMBL (ELIXIR-EBI); IMEx Consortium (ELIXIR-EBI and ELIXIR-IT), Biostudies (ELIXIR-EBI), Orphanet (ELIXIR-FR), RD-Connect portal (ELIXIR-ES), DisGeNET (ELIXIR-ES), eDGAR (ELIXIR-IT), Phyre (ELIXIR-UK), Genomic Hyperbrowser (ELIXIR-NO), PanDrugs (ELIXIR-ES), bio.tools (ELIXIR-DK), FAIRsharing.org (ELIXIR-UK), identifiers.org (ELIXIR-UK), WikiPathways, REACTOME, OpenEBench (ELIXIR Tools Platform), FAIR software and data stewardship services (ELIXIR Interoperability Platform) (ELIXIR-NL, ELIXIR-CZ, ELIXIR-UK, ELIXIR-IE), eLearning material and tools for ELIXIR services (ELIXIR Training Platform) (ELIXIR-UK, ELIXIR-SI and ELIXIR-IT).

WP1 (coordination) will support the RD community by aligning and interconnecting the available International RD infrastructure with the ELIXIR infrastructure and platforms and building upon the services provided by ELIXIR nodes ensuring alignment, collaboration and maximising the use of resources with international consortia such as the EJP-RD[5], GA4GH[6], BBMRI-ERIC[7], EOSC[8], IRDiRC, Orphanet[9], the GO FAIR implementation networks[10] and data generating projects such as Solve-RD[11].

WP2 (infrastructure) will investigate and coordinate an ELIXIR contribution to the development of an international RD-oriented data analysis environment, where available knowledge on RD (from the Orphanet database) in combination with data from different ELIXIR databases and resources, together with different genomic data analysis tools, interoperate to facilitate RD research. ELIXIR nodes have large experience in developing tools for genotyping, which is essential for the detection of RD causal mutations.

[1] http://www.irdirc.org/
[2] https://www.elixir-europe.org/about-us/how-funded/eu-projects/excelerate/wp8
[3] ELIXIR RD implementation studies
[4] https://www.elixir-europe.org/platforms/data/core-data-resources
[5] http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/sc1-bhc-04-2018.html
[6] https://www.ga4gh.org/
[7] http://www.bbmri-eric.eu/
[8] https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud
[9] https://www.orpha.net/consor/cgi-bin/index.php
[10] https://www.go-fair.org/implementation-networks/
[11] http://solve-rd.eu/

However, a systematic evaluation of ELIXIR tools tailored to answer the specific needs of the RD community (e.g. minimum depth of coverage, accuracy of genotype calls, etc.) has not been carried out to date for a proper primary data analysis step. To facilitate the systematic evaluation of tools and workflows by RD researchers, this proposal will work closely together with OpenEBench[12] (ELIXIR Tools Platform) to identify metrics aiming to evaluate those systems from a RD perspective. This effort will leverage already registered information for RD tools and workflows at bio.tools (ELIXIR Tools Platform) and genomic standards available in OpenEbench in the context of ELIXIR-EXCELERATE WP8 (catalogue RD resources[13] + EXCELERATE-Milestone 8.2[14]).

WP3 (interoperability) will investigate and coordinate the contribution of ELIXIR nodes to increasing capabilities for scaling up FAIRification of RD resources. Resources involved in WP2 will be investigated for their potential to be extended with FAIRification capabilities by ELIXIR services. Specific resources will be extended with such capabilities based on an evaluation of their potential for scaling up FAIRification, and the FAIRification needs for data used from these resources. We will build on the standard process to make RD data more FAIR compliant, the services that support FAIRification, and the standardized process to make RD genome data comparable developed in the context of the ELIXIR-EXCELERATE WP8 (Deliverable D8.2[15]). Our approach is to conduct an investigation of capabilities across ELIXIR nodes and coordinate collaboration between service providers.

WP4 will provide training of WP2 developments to enforce the end users, namely laboratory geneticists and RD researchers in their skills to use analysis tools, identify variants and targetable pathways for the better diagnosis and later treatment of RD patients.

Finally, all this proposal has been built around the combined activity of the ELIXIR RD community and the ELIXIR Interoperability, Compute, Tools and Training platform.

Description of work

WP1-Coordination and linkage with existing infrastructures, projects and initiatives	
Lead	Sergi Beltran (ES) and Marco Roos (NL)
Members	LU, ES, FR, HU, SI, ELIXIR RD-Community
Estimate	5 PMs
The international RD community is invested in working collaboratively across borders in order to bring diagnoses and therapies to patients, in alignment with IRDIRC vision of enabling all people living with a rare disease to receive an accurate diagnosis, care, and available therapy within one year of coming to medical attention delivering. During the past years and as envisioned in our roadmap, the ELIXIR Rare Disease Use Case -successfully transitioned into the ELIXIR Rare Disease Community- has been supporting the community by aligning and interconnecting international RD infrastructures with the general ELIXIR infrastructure and platforms and building upon the services provided by ELIXIR nodes. WP1 will work towards ensuring continuity of this work, supporting alignment, collaboration and maximising the use of resources with existing RD infrastructures, projects and initiatives. Also, this WP will work towards defining the ELIXIR RD Community future strategic plan and transitioning its coordination.	

[12] https://elixir.bsc.es/html/dashboard

[13] https://rare-diseases.bio.tools

[14] https://docs.google.com/presentation/d/1wXd3OpJtqHEtrlWS64Vaswl0d9U1V2-gtErVwnaWWdo/edit?usp=sharing

[15] https://docs.google.com/document/d/1T9pG9dLoxDM-bekv1ebNQD1ZZlFQWUCrytvDHWPP3DM/edit?usp=sharing

Task 1.1: Coordination of the overall ELIXIR RD Community (M1-M24, 2PM)
Leads: NL1, ES1 **. Participants:** LU1, ES2, FR1, HU1, SI1
The commitment of the ELIXIR RD community to contribute to the international RD infrastructure efforts for RD research requires to involve and coordinate many partners from different ELIXIR nodes, platforms and communities. Therefore, Task 1.1 will:

- Monitor and coordinate collaboration activities within all ELIXIR partners involved in this proposal and included in WP2 (infrastructure), WP3 (interoperability) and WP4 (training).
- Identify ongoing activities from ELIXIR nodes, platforms, the Human Data communities and other ELIXIR communities not partnering in this application that might be relevant for the RD community.
- Coordinate node contributions and collaboration between nodes to scale up FAIRification as defined in Task 3.1 (WP3-interoperability).

Task 1.2: Alignment with existing infrastructures, projects and initiatives (M1-M18, 2PMs)
Lead: ES1**. Participants:** ELIXIR RD-Community
Adoption of a global infrastructure for RD research by the international RD community, including European reference networks (ERNs), requires bundling capacity of ELIXIR with that in consortia such as the EJP-RD, GA4GH, BBMRI-ERIC, IRDiRC, Orphanet, the GO FAIR implementation networks and data generating projects such as Solve-RD. Current examples are Beacons, Phenopackets, implementation of data discovery APIs, privacy-preserving record linkage, and the implementation of FAIR principles in biobanks, registries and knowledge bases such as Orphanet. The success of this adoption also depends on the trust of stakeholders in long-term sustainability. Therefore, task 1.2 will:

- Map out and align activities included in the ELIXIR Rare Disease roadmap, the Infrastructure implementation study and the current community-led implementation study with ongoing international projects and initiatives. Mapping results will be used to define and execute specific contributions from WP2 (infrastructure), WP3 (interoperability) and WP4 (training) of this proposal.
- Identify, evaluate and support long-term sustainability of key RD focused European infrastructures and/or resources.

Task 1.3: Transition of the ELIXIR RD community coordination (M12-M24, 1PM)
Leads: NL1**, ES1. Participants:** ELIXIR RD-Community
Since its inception in 2015 and defined in the roadmap presented last year, the ELIXIR RD Use Case led by Marco Roos (ELIXIR-NL) and Ivo Gut (ELIXIR-ES) has been supporting RD researchers by (i) providing a global bioinformatics infrastructure for research, (ii) improving interoperability of data resources in terms of quality (FAIR principles) and standardization (ontologies) and (iii) providing training for the RD community. In this task we will promote discussion on future leadership of the ELIXIR RD community and organise the transition process if required.

WP2- Infrastructure for RD research	
Lead	Joaquin Dopazo (ES) and Marc Hanauer (FR)
Members	EBI, ES, FR, IT, NL, NO, UK, CH, SI, HU
Estimate	7 PMs

This WP aims to develop an RD-oriented data analysis environment, where available knowledge on RD (from the Orphanet database) in combination with data from different ELIXIR databases and resources, together with different genomic data analysis tools, interoperate to facilitate RD research. By means of

workshops the different contributors will F2F (or virtually) meet and present their resources and tools to identify existing and potential methods of interoperability. Uses case will be identified. The WP will identify what is achievable currently and what developments are required to increase interoperability among resources and tools.

The specific objectives of WP2 are:

1. Detect RD-related data available at several public general-purpose databases (ArrayExpress, ChEMBL, DisGeNET, eDGAR, UniProt, etc.) (Task 2.1)
2. Define RD-related disease maps from general purpose biological knowledge repositories (WikiPathways, REACTOME) and to investigate how to integrate them into an analytical and meta-analytical framework that include different analysis tools (Task 2.2)
3. Define how different types of tools for processing RD raw data (either new or from EGA) from genome resequencing experiments, must be specifically tailored to the necessities of the RD community (Tasks 2.3 and 2.4)
4. Define how tools or platforms for the analysis and interpretation of the results (genome resequencing, gene expression, enrichment, network and pathway analysis) can be integrated in an interoperable environment that will benefit research, disease gene discovery and drug repurposing in the field of RD. (Task 2.5)

Task 2.1: Extraction of RD-related information from databases and repositories
Leads: ES3, FR1**. Participants**: ES4, EBI1, FR1, IT3, IT4, IT6
Development of connectors that use Orphanet gene lists and tissue information in order to allow automatic queries to UniProt, ChEMBL, DisGeNET, eDGAR and PanDrugs to extract gene or drug data and information relevant for a chosen subset of RDs. Whenever is possible, these interoperability connection tools will make use of existent APIs in the databases. Some of these connections between resources can be done using federated queries to the RDF endpoints of UniProt, ChEMBL and DisGeNET, among others.

Task 2.2: Definition of RD-relevant pathways/sub-pathways
Lead: ES2**. Participants:** ES3, IT4, NL2, NL3, FR1, EBI2
Orphanet information will be used to define disease pathways within WikiPathways and REACTOME that are relevant for a chosen subset of RDs. Such subpathways can be directly included in the corresponding tools of Task 2.5 or, alternatively, the tools can be directly invoked from the repositories.

Task 2.3: Tools for data processing
Lead: IT1. **Participants:** ES2, ES4, FR2, NO1, NO2, NO3, SI1, HU1
Identifying relevant standard analysis workflows and benchmarks to provide guidelines and best practices for targeted resequencing experiments. Identify requirements for making tools for primary data processing interoperable with analysis tools in WP3 (interoperability).

Task 2.4: Tools for data annotation
Lead: UK1. **Participants**: ES1, ES4 , FR2, IT3, IT4, IT7, IT8, UK2 ELIXIR-UK
There are numerous resources and databases that can enrich the annotation of RD including sequence, structure, function and interactions. This task will aim to assess how interoperable current resources are and how they can be improved. This will be assessed via some chosen test cases.

Task 2.5: Tools for data analysis and interpretation
Lead: ES2. **Participants**: ES1, ES3, ES4, ES5, FR2, IT3, IT4, IT5, IT6, NO1, NO2, NO3, HU1, HU2, EBI2, NL2, NL3, SI1
Definition of a subset of gene expression, pathway or network analysis tools (initially, Hipathia, DisGeNET, eDGAR, Genomic HyperBrowser, PathVisio) that can be used to analyze RD 'omic data. Identification of the requirements to build an infrastructure of connectors of these tools among them and with the

databases (ArrayExpress, ChEMBL), knowledge repositories (DisGeNET, eDGAR, Uniprot, PanDrugs) and pathways (WikiPathways, REACTOME) and demonstration of the feasibility of this interoperability with a subset of chosen cases.

WP3- Interoperability of RD resources	
Lead	Marco Roos (NL), Dietlind Gerloff (LU)
Members	CZ, IT, LU, NL, NO, PT, SI, ES
Estimate	4 PMs

Implementation of FAIR principles will continue to be a prime target for accelerating research on rare disease data, given the sensitivity and distribution of rare disease data that require the ability to efficiently analyse sparse data across rare disease resources. The aim of this WP is to increase capabilities for scaling up FAIRification in the rare disease community through ELIXIR FAIRification services. This can be achieved by adding FAIRification capabilities to widely used tools, such as by deployment and adaptation of services that support the FAIRification workflow (EXCELERATE Deliverable D8.2). Our approach is to conduct an investigation of capabilities across ELIXIR nodes and coordinate collaboration between service providers. We build on outcomes of previous work for this investigation: current ELIXIR node activities reported in the recent survey for supporting FAIR in the rare disease community, and the resources listed in WP2 as potential targets for adding 'FAIRification-for-purpose' capabilities. The work will be carried out as a combined activity of the rare disease community and the ELIXIR Interoperability platform, with strong links to the Compute platform and the Tools platform.

Task 3.1 ***Investigation of services to scale up FAIRification*** **(M1-M6, 1PM)**
Lead: LU1. **Participants**: CZ, IT , NL, NO, PT, SI, ES
In this task, we identify what ELIXIR nodes can do to contribute to scaling up FAIRification in the rare disease community in general, and specifically in relation to the activities targeted in WP2: annotation and analysis of genomic data for interpretation in the context of disease-related pathways. The investigation will cover (i) existing services provided at ELIXIR nodes (using the RD-community survey across all ELIXIR nodes from Summer 2018 as a starting point), (ii) ELIXIR platform roadmaps, and (iii) other rare disease projects that also support FAIR (particularly the EJPRD). The purposes of the FAIRification that is facilitated by the services will also be charted. The resources named as the initial set for WP2 (infrastructure) are of particular interest here: analysing the potential for extending them with 'FAIR-for-purpose' capabilities is facilitated through the alignment between WP2 (infrastructure) and WP3 (interoperability) and with the participating service providers. The investigation will result in a plan for coordinating existing node activities towards an effective contribution to scaling up FAIRification in the rare disease community (Task 1.1), and prioritisation suggestions for RD-relevant resources that can be targeted in Task 3.2.

Task 3.2 ***Implementation of services to scale up FAIRification*** **(M7-M24, 3PM)**
Lead: NL2**. Participants**: LU2
In this task, we will evaluate in-depth for at least one resource how to extend it with 'FAIRification-for-purpose' capabilities, following the identification of potential targets in Task 3.1. We build on the recommendations resulting from the investigation in Task 3.1, such to ensure that rare disease

FAIRification services are optimally aligned with the roadmaps of the interoperability, compute and tools platforms.

WP4 - Training and Outreach	
Lead	Brane Leskošek (SI), Balint L. Balint (HU)
Members	ES, SE, IT, LU, NL, DE, HU, SI, PT, NO
Estimate	4 PMs

In the last two years WGS of large clinical sample cohorts became a reality, thus the RD community can disseminate the acquired knowledge towards the larger clinical geneticist expert group. This work package aims to provide RD-specific training for end users, the clinical laboratory geneticist newly engaged in the analysis of RD data or in the process of adapting their procedures to new technological opportunities and/or new legal/ethical standards. These trainings will rely on the RD tools used in RD infrastructure (WP2).

A set of planned developments listed in WP2 (infrastructure) aim to enforce the end users, namely laboratory geneticists and RD researchers in their knowledge, skills and ability to use analysis tools, identify genetic variants and targetable pathways for the better understanding of RS phenotypes, faster diagnosis and later novel treatments for RD patients.

The activities planned in WP2 (infrastructure) cover the “workflows-data analysis-data interpretation” trio and are focusing on the development of standard workflows with optimal interoperability capabilities (Task 2.3), demonstration of interoperability of the collection of tools and databases (Task 2.5).
Our planned training activities will mirror the “workflows-data analysis-data interpretation” trio and will be based on the surveys and gap analysis completed previously within the RD community (2017) about the available training materials. These surveys summarized a wide range of tools and training materials in the area of data storage, data security, multi-omics integration, standard analysis workflows.

In order to take advantage of the widening of the RD data resources, tools and standards, training activities will be provided for clinical geneticists with the focus on maximizing interoperability with the available and newly developed tools within the ELIXIR community like the local EGA tools, FHD infrastructures and tools with GUI like Galaxy. Outreach and visibility for the RD community will be implemented via presentations at RD and clinical genetics conferences/meetings and announcements through the ELIXIR Training Portal TeSS. Quality assessment of the training activities will be monitored by the standardized tools of the ELIXIR Training Platform namely short term and long term feedback questionnaires through implementation of the Training Toolkit.

Task 4.1: Training guideline development for tools for data processing in alignment with Task 2.3
Lead: HU1**. Participants**: IT1, DE1, PT1, LU3, IT8, SI1
Identification and prioritization of the training resources required in order to provide specific training to laboratory geneticists, newly engaged in the analysis of RD data or in the process of adapting their procedures to new technological opportunities and/or new legal/ethical standards. Training guidelines will be developed after evaluation of training materials in a selection of RD use cases. We will prototype training materials and test them during F2F meetings with the previously defined end users. Based on the evaluation of the test cases guidelines for further training material developments will be generated.

Special focus will be given to emphasize the benefits of integrating available workflows into the novel ELIXIR developments like local EGA. This subtask will deal with the activities proposed in WP 2.3, and in particular with the tools and workflows for the analysis of the data that will be therein developed as well as the added value of the local EGA and FHD developments.

Task 4.2:Training guideline development for tools for data analysis and interpretation in alignment with Task 2.5
Lead: SI1. **Participants**: DE1, HU1, HU2, ES2, PT1, IT3, IT4, NL2, NL3, SE1, NO1, NO2, IT8
By using the method described at Task 4.1 we will perform the identification of training resources needed for providing training for laboratory geneticists in large field of available tools and resources of data analysis and interpretation. As a test case we will focus on a chosen subset of tools that will be selected as best compliant with best practices of interoperability in Task 2.5. Evaluation of test cases will be used for finalizing the Guidelines for further developments of training materials.

Alignment with Evaluation Criteria

This proposal meet the different criteria of evaluation. Detailed information concerning each of the five high-level Criteria can be found in Annex I (below).

Expected outcome

please see Delivery Schedule - by Node

Plan for disseminating outcome

The project progress will be communicated annually at the ELIXIR All Hands meeting and in other presentations whenever appropriate. The outcomes of each WP or Task will be communicated and demonstrated, if appropriate, in the ELIXIR webinar series (organised by the ELIXIR Hub in coordination with the WP leads).

Statement that the proposed work is not being supported by alternative funding sources

This proposal is not being supported by any alternative funding sources and efforts will be made in WP1 to ensure alignment with other rare disease infrastructures, projects and consortia to avoid double funding.

Project Timeline:

Timeline of the project is provided in the following Gantt chart

Delivery Schedule - by Node:

Milestones:	**Deliverables:**
1. **M1.1** Identification of activities relevant for RD research across ELIXIR nodes, platforms and communities (Task1.1, M6; LU, ES, NL)	1. **D1.1** Mapping results of the ELIXIR RD community activities with main international projects and initiatives (Task1.2, M18, ES, NL, LU)

2. **M1.2** Identification of activities relevant for the ELIXIR RD community across EJP-RD, Solve-RD and GA4GH (Task1.2, M12; ES) 3. **M2.1**. DisGeNET, eDGAR, PanDrugs with information mapped onto RD (Orphanet) identifiers (M9, ES, IT) 4. **M2.2**. Definition of a set RD-related pathways for subsequent cases of use and specifications to define RD-related pathways in REACTOME and WikiPathways (M12, FR1, ES2, NL2, NL3) 5. **M2.3.** Test case identification of relevant standard analysis and workflows for targeted resequencing experiments (M12, IT1) 6. **M2.4.** Identification of potential methods of interoperability among resources and databases that can enrich the annotation of RD with structure, function and interaction information (M12, UK1) 7. **M2.5**. Test implementation of the tools (Hipathia, DisGeNET, eDGAR, Genomic Hyperbrowser, PathVisio, PanDrugs) tool for pathway analysis interoperable with ArrayExpress, UniProt, REACTOME and WikiPathways in a case use subset or RDs (M24, ES2, ES3, ES5, IT3, IT4, NO1, NO2, NL2, NL3, EBI1, EBI2) 8. **M3.1** List of activities to coordinate between nodes have been defined (Task 3.1, M4; LU) 9. **M3.2** Target resource(s) to implement or extend FAIRification capabilities, the purpose the resulting FAIR data will serve, and service(s) to implement these capabilities have been defined (Task 3.1, M6, LU, NL) 10. **M3.3** A service architecture for extending a resource with FAIRification capabilities has been designed (Task 3.2, M10, NL, LU) 11. **M4.1** Annual f2f workshop (together with ELIXIR AHM or similar event and coordinated with RD, EGA or FHD and Galaxy communities) (M12, M24)(HU) 12. **M4.2** Training f2f workshops in alignment with WP2 needs (M10, M16, M22)(SI) 13. **M4.3** All Short term surveys collected and Long term surveys sent to all participants (M24)(HU)	2. **D2.1.** Specifications to enable DisGenet, eDGAR and PanDrugs – Orphanet mapping (M9; ES, IT) 3. **D2.2**. Definition of a subset of RDs to extract RD-related pathways for subsequent cases of use and specifications to define RD-related pathways in REACTOME and WikiPathways (M12; FR, ES, NL) 4. **D2.3.** Report on standard analysis and workflows for targeted resequencing experiments (M12, IT) 5. **D2.4.** Report on the interoperability among resources and databases that enrich RD annotation with structure, function and interaction information (M12, UK) 6. **D2.5**. Report on interoperability of tools (Hipathia, DisGeNET, eDGAR, Genomic Hyperbrowser, PanDrugs) with resources (ArrayExpress, UniProt, REACTOME and WikiPathways) in a case use subset or RDs (M24, ES, IT, NO) 7. **D3.1** Report on the status quo regarding FAIR capabilities of ELIXIR node resources (data, services, purpose) of particular interest to the RD community, with suggested example(s) for extension in Task 3.2 (Task 3.1, M6, LU, NL) 8. **D3.2** Prototype implementation of extended FAIRification capabilities in selected resource. (Task 3.2, M24, NL, LU) 9. **D4.1** Report on existing gap analysis, training needs and available training resources of the RD community (M6) (SI, HU) 10. **D4.2** Training resources and guidelines for further developments of RD specific training materials and outreach in alignment with WP2 activities (M12, updates M18, M24) (HU, SI) 11. **D4.3** Pilot online (e-learning) training resources extracted from f2f workshops and accessible via TeSS (M24) (SI)

Annex 1

Detailed alignment with Evaluation Criteria

This proposal meet the Evaluation Criteria described in the Request for Proposals. Detailed information concerning each of the different criteria is detailed below

1. **Scientific focus, scope, need**
   - Involvement of experts on developing bioinformatics resources and data analysis tools from ELIXIR nodes from different communities (human data communities) and platforms (interoperability, tools and training).
   - Contribution to the development of an international infrastructure for RD research and specific objectives towards this goal have been clearly defined and aligned with RD-Community needs.
   - ensure interconnection with international initiatives such as EJP-RD, GA4GH, EOSC, etc. (WP1-coordination)
   - Improvement of the quality of ELIXIR platform services and resources thanks to the work performed in WP2 (infrastructure) and WP3 (interoperability)
2. **Community served**
   - Contribution to the development of an international RD-oriented data analysis environment, where available knowledge on RD (from the Orphanet database) in combination with data from different ELIXIR databases and resources, together with different genomic data analysis tools, interoperate to facilitate RD research
   - Build upon previous implementation studies and the work done within the ELIXIR-EXCELERATE project
   - Outcomes will be disseminated to the ELIXIR community through the organisation of a webinar and the communication at the ELIXIR All Hands meeting and to the whole RD community through WP4 (training) that will provide training of WP2 (infrastructure) developments to enforce the end users in their skills to use analysis tools for the better diagnosis and later treatment of RD patients
3. **Quality of Science**
   - focus on improving RD resources, standards and ontologies and involves a whole WP (WP3-interoperability) to ensure improvements towards applying FAIR Data Principles.
   - enable interoperability of resources at different levels (robust APIs, use of standards and identifiers or metadata description) enabling RD research to analyse their data in a more efficient way.
4. **Towards supporting the mission of ELIXIR**
   - aligns with ELIXIR mission to coordinate ELIXIR resources so that they form a single infrastructure
   - coordinate, integrate and improve quality and interoperability of resources useful for RD research into an international rare disease framework
   - accelerate RD research towards bringing the infrastructure to analyse RD patient data and for treatment/ drug discovery
   - strengthen coordination and integration between the rare disease community and ELIXIR platforms building on previous work (implementation studies and EXCELERATE) mainly on FAIR services with the interoperability platform, RD resources with the tools and compute platform and on RD training activities with the training platform
5. **Sustainability and impact of the implementation**
   - Most of the resources and services addressed in this proposal are already included in ELIXIR national nodes and platform strategic roadmaps and are already planned to be sustained in the following years
   - By bringing those resources to the RD infrastructure we will improve their interoperability and increase their user access which will be, in return, beneficial for their long term sustainability

- impact of this implementation study will be measurable through the number of users that access the different resources involved in the proposal and their adoption by the rare disease community thanks to the alignment with RD projects and initiatives