EXCELERATE Deliverable 8.1

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: Portfolio of ELIXIR data resources and tools for the rare diseases communities	
WP No.	8	
Lead Beneficiary:	7 : CNIO	
WP Title	Use Case C: ELIXIR infrastructure for Rare Disease research	
Contractual delivery date:	31 August 2017	
Actual delivery date:	31 August 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:

Sergi Beltran, Eivind Coward, Leslie Matalonga, Finn Drablos, Hedi Peterson, Ivo Gut, Marco Roos.

Table of content

1. Executive Summary

There is a wide range of data resources and analysis methods used in the rare-disease area. Deliverable 8.1 report the efforts undertaken from the ELIXIR-EXCELERATE WP8 rare disease use case to create a dynamic portfolio of ELIXIR data resources and analysis tools useful to the rare disease communities.

Since the start of the project, we have reviewed the ELIXIR-EXCELERATE current data resources and evaluated their usability and potential impact on the rare disease community. One critical aspect of the development of the registry has been to engage the different communities in the submission and rating of the tools. For this task, we worked together with representatives of the major projects in the field of rare diseases to create a customized catalogue of ELIXIR tools and services devoted to assist them in the diagnosis of rare disease patients and the development of new therapies, which are the main goals established by the International Rare Diseases Research Consortium (http://www.irdirc.org).

Through the elaboration and dissemination of a rare disease resource survey, we have started to deploy of a link between the end users and the tools developers that 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.

To publish and make available to the rare disease community our catalogue of resources, we have used the ELIXIR registry (https://bio.tools/) in collaboration with ELIXIR- EXCELERATE WP1 and WP2. ELIXIR registry is intended to be a reference for the research community, as it will reflect the quality and the real-time status of the services included on it. Displaying our catalogue of rare disease resources into the ELIXIR registry environment allow users from the different countries, communities and projects to discover which are the tools available for rare disease research at a given time and provide them with instructions for their correct use.

Deliverable 8.1: “Portfolio of ELIXIR data resources and tools for the rare diseases communities” defines the outcome of this work.

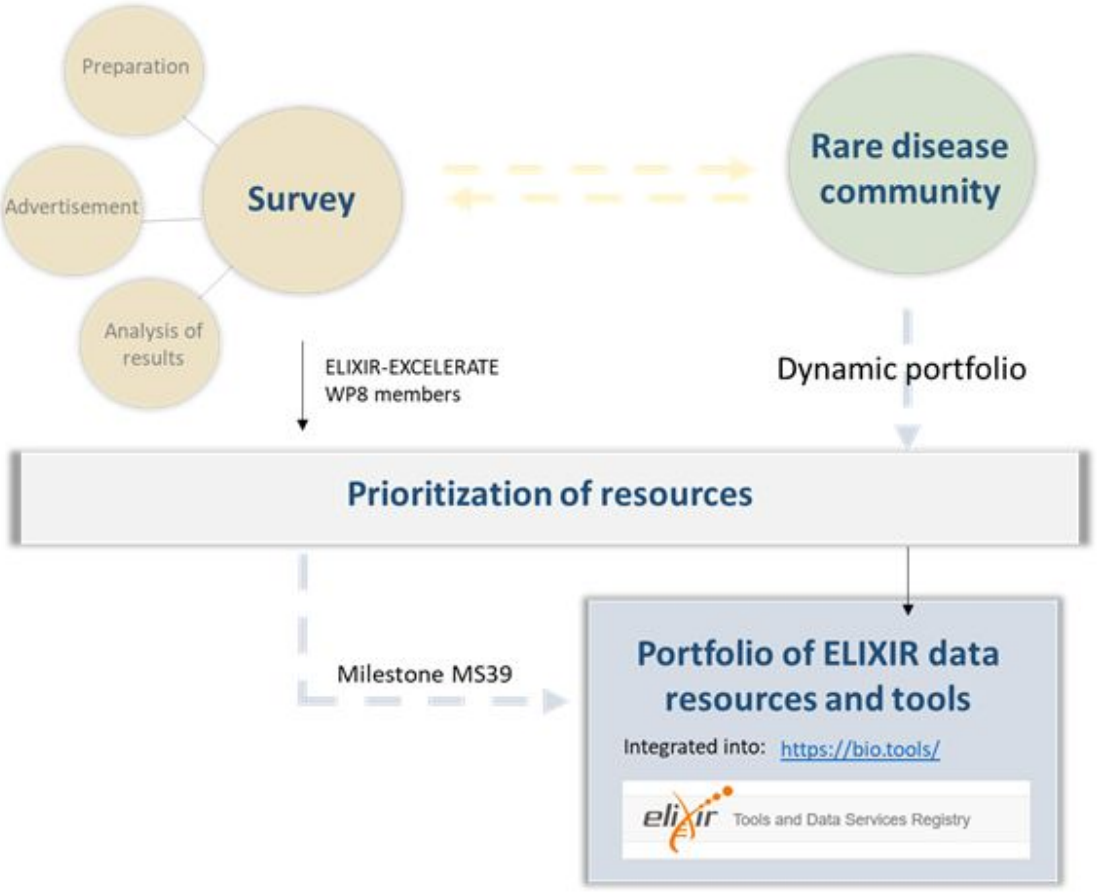


Figure 1. Schema of the elaboration of the rare disease portfolio of data resources and tools.

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

N/A

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		
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 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 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

6. Appendix 1: REPORT. Portfolio of ELIXIR data resources and tools for the rare diseases communitiestle

Summary

There is a wide range of data resources and analysis methods used in the rare-disease area. Deliverable 8.1 report the efforts undertaken from the ELIXIR-EXCELERATE WP8 rare disease use case to create a dynamic portfolio of ELIXIR data resources and analysis tools useful to the rare disease communities.

Since the start of the project, we have reviewed the ELIXIR-EXCELERATE current data resources and evaluated their usability and potential impact on the rare disease community. One critical aspect of the development of the registry has been to engage the different communities in the submission and rating of the tools. For this task, we worked together with representatives of the major projects in the field of rare diseases to create a customized catalogue of ELIXIR tools and services devoted to assist them in the diagnosis of rare disease patients and the development of new therapies, which are the main goals established by the International Rare Diseases Research Consortium (http://www.irdirc.org).

Through the elaboration and dissemination of a rare disease resource survey, we have started to deploy of a link between the end users and the tools developers that 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.

To publish and make available to the rare disease community our catalogue of resources, we have used the ELIXIR registry (https://bio.tools/) in collaboration with

ELIXIR- EXCELERATE WP1 and WP2. ELIXIR registry is intended to be a reference for the research community, as it will reflect the quality and the real-time status of the services included on it. Displaying our catalogue of rare disease resources into the ELIXIR registry environment allow users from the different countries, communities and projects to discover which are the tools available for rare disease research at a given time and provide them with instructions for their correct use.

Deliverable 8.1: "Portfolio of ELIXIR data resources and tools for the rare diseases communities" defines the outcome of this work.

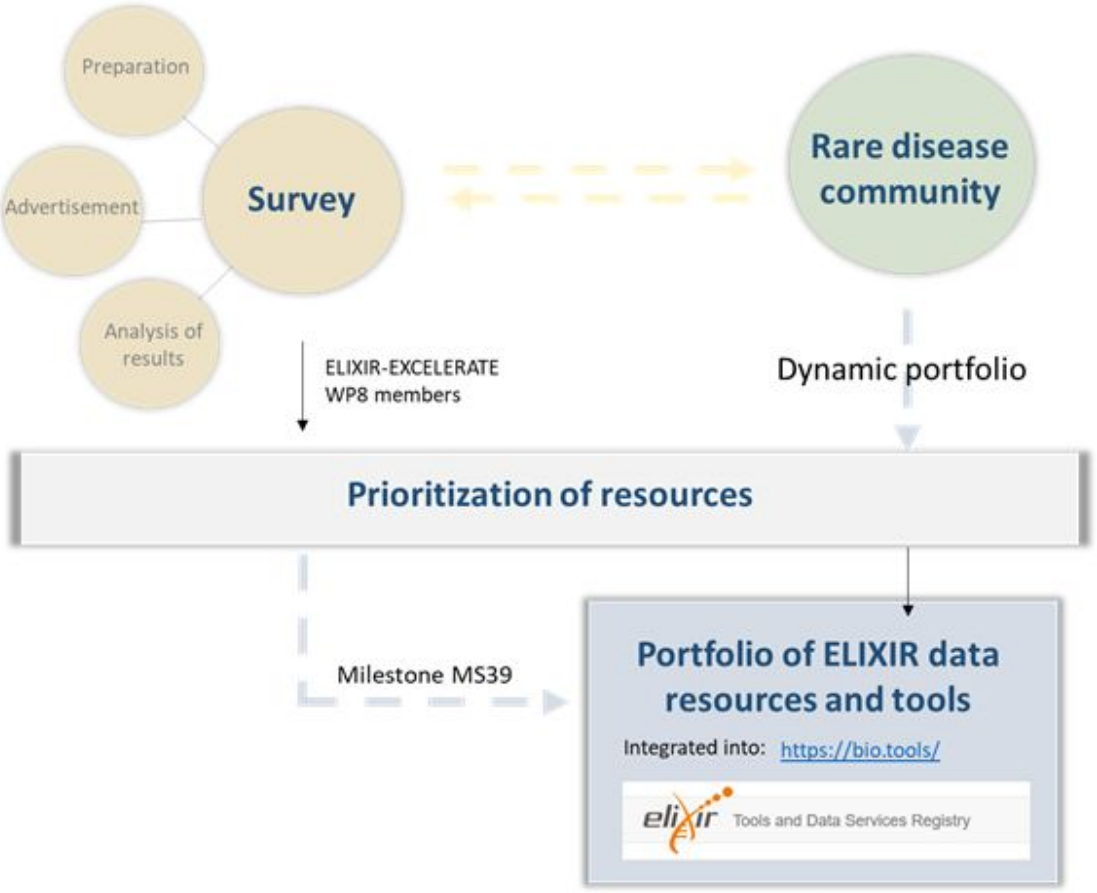


Figure 1. Schema of the elaboration of the rare disease portfolio of data resources and tools.

1.- Rare diseases data resources and tools survey

The survey has been designed as a part of Subtask 8.1.1 of ELIXIR-EXCELERATE WP8, "Monitoring of resources and tools" for Rare Diseases. The aim was to determine the current usage and needs of the rare disease (RD) community. It helped in

prioritizing the most important bioinformatics tools, determine the needs to be improved, and ultimately will lead to better ELIXIR services adapted to the requirements of the community.

1.1 Design and advertisement

The survey was designed using Google Forms, after an extensive discussion with contributions from several WP8 members and RD liaisons.

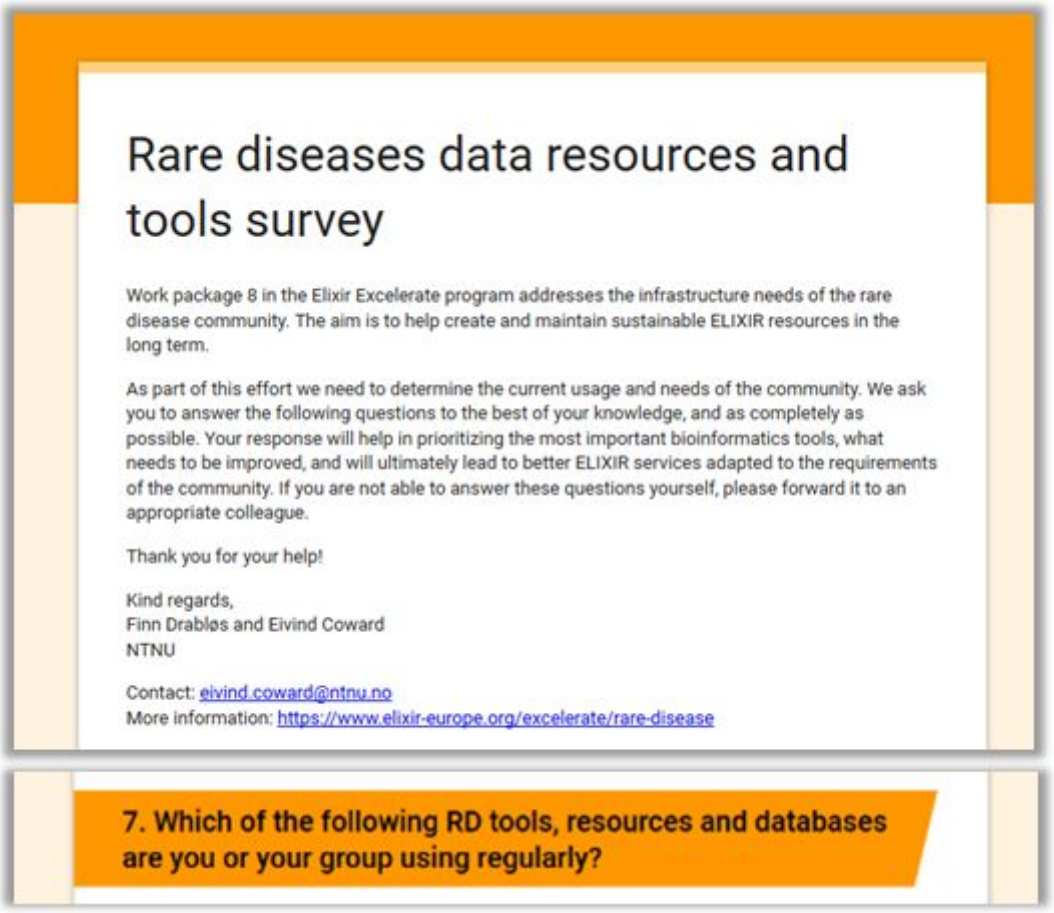


Figure 2. Rare diseases data resources and tools survey: header and example of the addressed questions. Full survey can be found here.

The survey was advertised to the community using established networks, such as RD-Connect (http://rd-connect.eu/), Neuromics (http://rd-neuromics.eu/), Eurenomics (https://eurenomics.eu/), Orphanet (http://www.orpha.net/consor/cgi-bin/index.php), and the ELIXIR WP8 itself. Moreover, members of the WP8 group distributed the call to colleagues in their respective countries. In addition, external key persons for specific

domains were asked to complete the survey and forward the request. These key persons were Claudio Carta (disease registries), Virginie Bros-Facer (patient representative), Mary Wang (biobanks), and Jérôme Weinbach (RaDiCo network).

1.2 Response and results

The first deadline was set to November 7, 2016. Because of limited response, the deadline was extended to January 16, 2017 and by this date, 47 responses have been received.

The geographical distribution is quite skewed, with more than a half coming from the Netherlands. The profiles of the respondents are fairly evenly distributed between bioinformaticians, wet lab researchers, clinicians, and "others" (mostly describing themselves as combinations of the above or other types of researchers and academics).

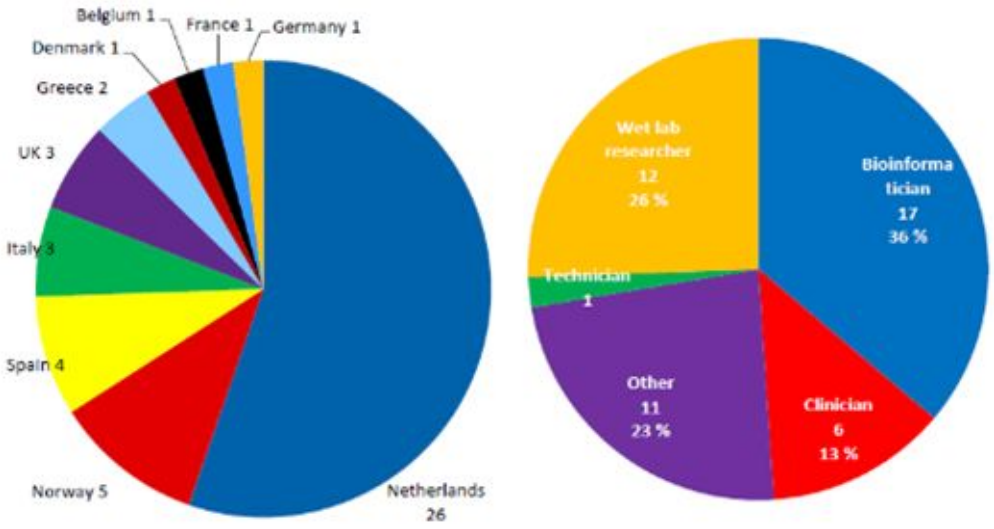
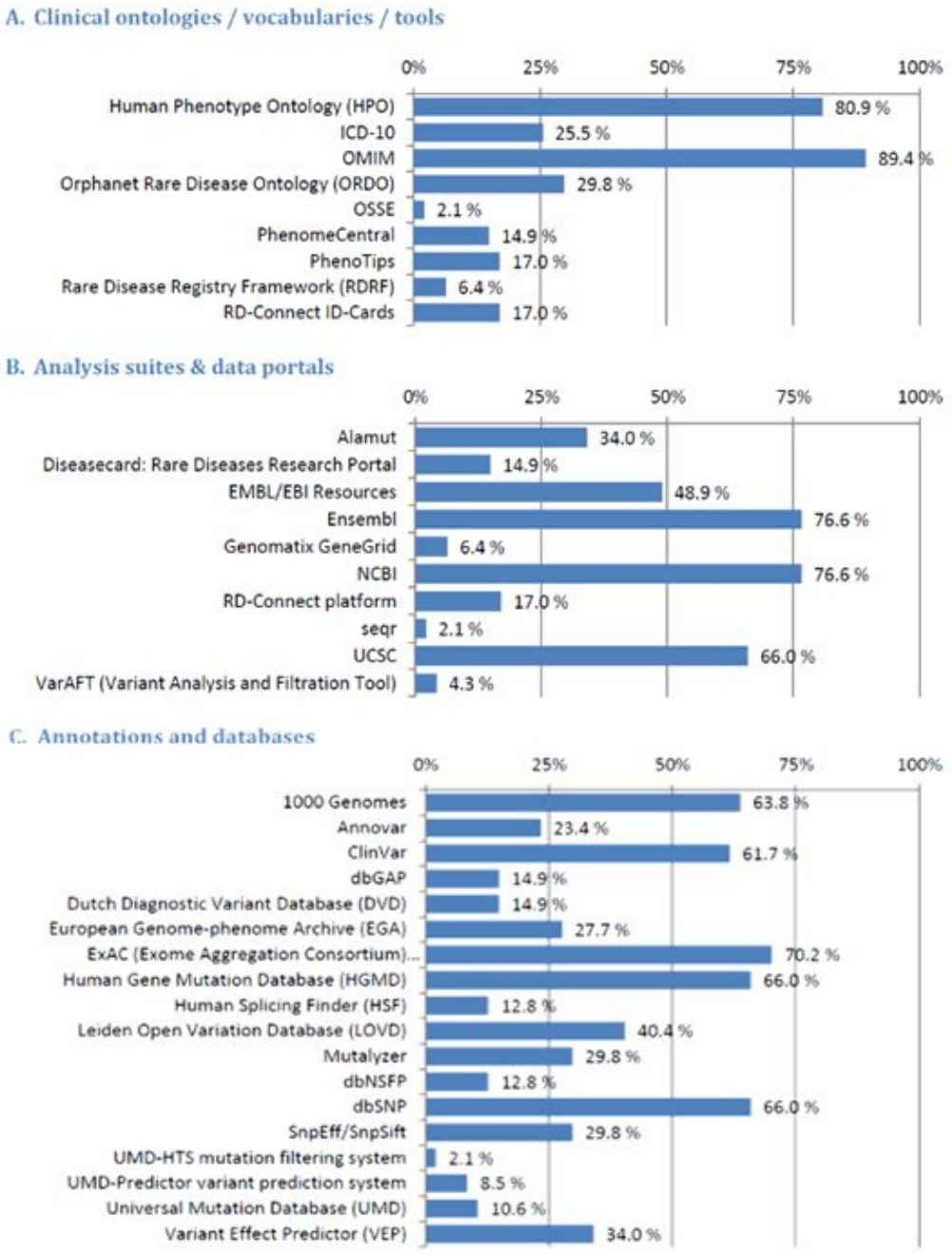


Figure 3. Geographical distribution and profile of survey respondents.

1.2.1 Current usage of tools and resources

Question 7 constitutes a major first part of the survey: *Which of the following RD tools, resources and databases are you or your group using regularly?* An extensive checkbox list with tools and resources in five different categories was presented. The user was able to select any number of alternatives in each category.

In each case, the result is presented as a percentage of the total respondent body, even if some of them ticked no alternatives in some of the categories.



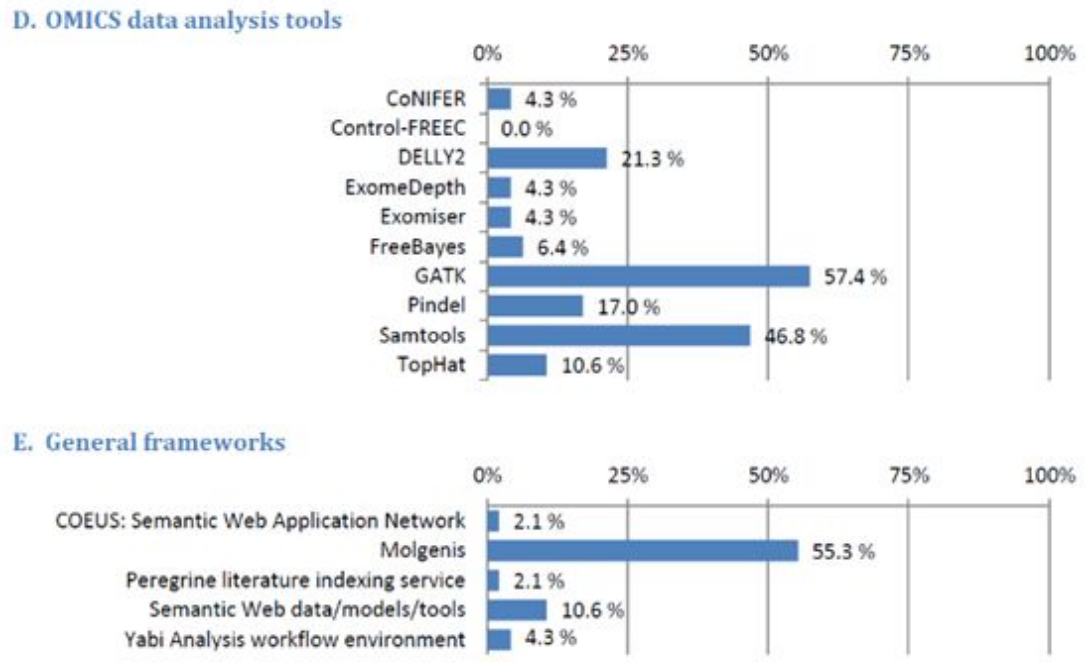


Figure 4: tools, resources and databases used regularly by the respondent core.

Other tools and resources:

For tools and resources not in the list, there was an open question 8: *Which other tools or databases are you regularly using in your RD research?*

- 21 of the respondents added in total 36 different resources in this field.

1.2.2 Need for improvement and development

Respondents were asked to rate the tools in four different categories. The rating is general, applying to all the tools. This is a limitation, but it was regarded as necessary partly for technical reasons, and partly because rating each tool would have made the survey quite time demanding for the user. However, the respondents were urged to add specific comments to tools in the next question.

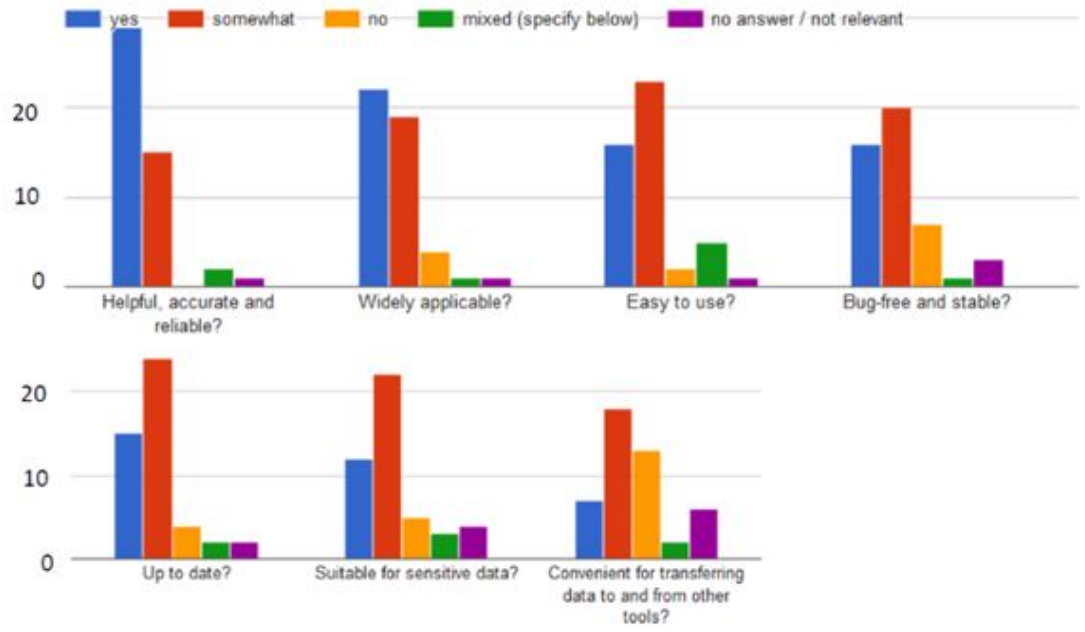


Figure 5: Rating of tools by the respondent core.

Summing up the comments on the different tools (see full results), there is a demand for improvement and development of a diverse range of tools. CNV tools and phenotypic information in databases are also requested by several users.

The analysis and results of this survey have also been presented during the ELIXIR All Hands 2017 (21-23 March 2017 in the Barceló Aran Mantegna, Rome, Italy). Poster nº10 presented by Eivind Coward.

2.- Catalogue of Rare diseases data resources and tools

The results of the survey presented above allowed us to establish a first prioritisation of the different resources and to publish a catalogue of resources, data sources and methods assessed for the RD communities.

On our first release (February 2017), the catalogue comprised around fifty resources and analysis tools (Table 1). This part of the work enabled us to reach milestone MS39 (MS8.1) concerning the publication of a catalogue of resources, data sources and methods assessed in collaboration with the rare-diseases communities. MS39 has been achieved and reported at the end of February 2017. The report on this milestone can be found here.

Table 1. List of tools prioritized after analyzing the survey results.
Last update: 17 July 2017

Priority°	Resource	Type of data	PMID*	Present in Biotools	Added in biotools
High	SAMtools	Tool	19505943	yes	January 2017
High	DELLY2	Tool	22962449	yes	February 2017
High	GATK	Tool	20644199	yes	January 2017
High	Molgenis	Tool	27153686	yes	2015
High	Variant Effect Predictor (VEP)	Tool	27268795	yes	January 2017
High	SnpEff	Tool	22728672	yes	January 2017
High	dbSNP	database	11125122	yes	January 2017
High	Mutalyzer	Tool	18000842	yes	February 2017
High	Leiden Open Variation Database (LOVD)	Tool	21520333	yes	January 2017
High	Human Gene Mutation Database (HGMD)	database	24077912	yes	February 2017
High	Exome aggregation consortium (ExAC)	database	27899611	yes	February 2017
High	European Genome-Phenome archive (EGA)	repository	26111507	yes	2015
High	ClinVar	database	24234437	yes	July 2017
High	1000 Genomes	database	26432245	yes	July 2017
High	UCSC	database portal	26590259	yes	2016
High	NCBI	database /some tools	https://www.ncbi.nlm.nih.gov/	yes	different entries
High	Ensembl	database	27899575	yes	different entries
High	EMBL/ EBI resources	database/tools associated	25845596	yes	2015

High	Alamut	database/tools associated	http://www.interactive-biosoftware.com/products/	yes	February 2017
High	Human Phenotype Ontology (HPO)	ontology	24217912	yes	February 2017
High	Online Mendelian Inheritance in Man (OMIM)	database	11752252	yes	2016
High	Orphanet Rare Disease Ontology (ORDO)	ontology	http://www.orphadata.org/cgi-bin/inc/ordo_orphanet.inc.php#	yes	2015
Moderate	International Classification of Diseases (ICD-10)	Ontology	25122944	yes	February 2017
Low	OSSE (An Online Sample Size Estimator)	Tool	http://osse.bii.a-star.edu.sg/references.php	yes	March 2017
Moderate	Phenome central	Database / repository	26251998	yes	March 2017
Moderate	Phenotips	Tool / registry	23636887	yes	February 2017
Moderate	Rare disease Registry Framework (RDRF)	Tool / registry	24982690	yes	February 2017
Moderate	RD-connect ID-Cards	database /registry	25029978	yes	March 2017
Moderate	Diseasecard: Rare disease research portal	database	23973272	yes	February 2017
Moderate	Genomatix GeneGrid	tool	https://genegrid.genomatix.com/grid/home	yes	February 2017
Moderate	RD-connect platform	tool	25029978	yes	February 2017
Low	seqr	tool	https://seqr.broadinstitute.org/	yes	March 2017
Moderate	Variant Analysis and Filtration Tool (VarAFT)	tool	http://varaft.eu (manuscript in preparation)	yes	February 2017
Moderate	Annovar	tool	20601685	yes	January 2017
Moderate	Database of Genotypes and Phenotypes (dbGAP)	database	24297256	yes	February 2017
Moderate	Dutch diagnostic Variant Database (DVD)	database /registry	https://trac.nbic.nl/dvd/	yes	February 2017
Moderate	Human Splicing Finder	tool	19339519	yes	February 2017
Moderate	dbNSFP	database	26555599	yes	February 2017

low	UMD-HTS mutation filtering system	server is temporarily unable			
Moderate	Universal Mutation Database (UMD) predictor variant predictor system	Tool	26842889	yes	February 2017
Moderate	Universal Mutation Database	database with links to dif tools	10612827	no	
Low	Exomedepth	tool	22942019	yes	February 2017
Low	Exomiser	tool	26562621	yes	February 2017
Moderate	FreeBayes	tool	https://github.com/ekg/freebayes	yes	January 2017
Moderate	Pindel	tool	19561018	yes	January 2017
Moderate	TopHat	tool	19289445	yes	January 2017
Moderate	Semantic web application network (COEUS)	database	23244467	yes	February 2017
low	Peregrine literature indexing service	tool	https://trac.nbic.nl/data-mining/	yes	February 2017
Moderate	Semantic Web data/models/tools	database/ tool	https://www.w3.org/2001/sw/wiki/Main_Page	no	
low	Yabi analysis workflow environment	tool	22333270	yes	February 2017
Moderate	SNPs and GO	tool	19514061	yes	2015
Moderate	eDGAR (a database of Disease-Gene Associations with annotated Relationships among genes)	database	http://edgar.biocomp.unibo.it	yes	March 2017

º High = more than 30% of the total respondent body, Moderate = between 5-30% of the total respondent body, Low= less of 5% of the total respondent body
* or website if PMID not available

The different tools were given different priorities to be included in the catalogue depending on the usefulness reported by survey respondents. A high priority was given to those tools reported as useful for RD research by more than 30% of the respondents. The priorities are used for planning and coordination; based on the survey we do not claim that they reflect the priorities of the complete RD community without bias.

Rare disease research tools and resources is a highly dynamic field, therefore we envision to constantly update our catalogue: adding new tools and removing the obsolete ones for the community. The dynamic catalogue is available here and will be updated in bio.tools periodically. For example, we have started a collaboration with the RD-connect project (http://rd-connect.eu/) for adding rare disease resources used in the context of the project to our catalogue of resources.
We are also working in collaboration with the newly established Joint Research Unit ELIXIR-IIB (Italian Infrastructure for Bioinformatics) to integrate Italian resources within the ELIXIR-EXCELERATE Rare Disease Use Case through an implementation study that will start next October 2017: "Implementation study for the integration of ELIXIR-IIB in ELIXIR Rare Diseases activities". Within this study 8 Italian resources will be integrated to the RD portfolio.

To date, more than 90 tools have been added to our dynamic RD resources catalogue.

3.- Integration of the catalogue of Rare diseases data resources and tools into ELIXIR bio.tools environment

We have elaborated a catalogue of RD resources which is available to the RD community through the ELIXIR: tools and data registry platform (https://bio.tools/). This part of the work has been done in collaboration with the ELIXIR tools platform, WP1: Tools Interoperability and Service Registry, led by Søren Brunak (DK) and Alfonso Valencia (ES). For this part of the project, we have worked in close collaboration with ELIXIR Estonia (liaison Hedi Peterson) to add resources useful for the RD community and previously prioritized by our WP8, to bio.tools.

From the first analysis, we prioritised 49 tools and resources to be included in bio.tools. 18 of these tools were already available in the system and we made new entries for 31 of them. All the information and resource descriptions have been curated by WP8

members. To date, more than 90 tools have been added to our dynamic RD resources catalogue, prioritized and integrated into bio.tools.

This portfolio of resources, available through ELIXIR bio.tools, will allow users from the different countries, communities and projects to discover which are the tools available for rare disease research at a given time. We have settle up different use cases to show the utility and benefits of the portfolio.

3.1 Use Cases

3.1.1: Search for a specific tool

The RD researcher / end user is interested in gathering information of a specific tool. The RD researcher will access bio.tools (https://bio.tools/) and type the name of the resource of interest. The ELIXIR bio.tools application will return a small descritption of the resource, a direct link to the website and a link with additional information, references and possible other actions (Figure 6).
Within the framework of the ELIXIR-EXCELERATE project (WP1 and WP2), additional information such as tool version and benchmarking metrics (performance, impact, usability, etc) is also envisioned to be displayed.

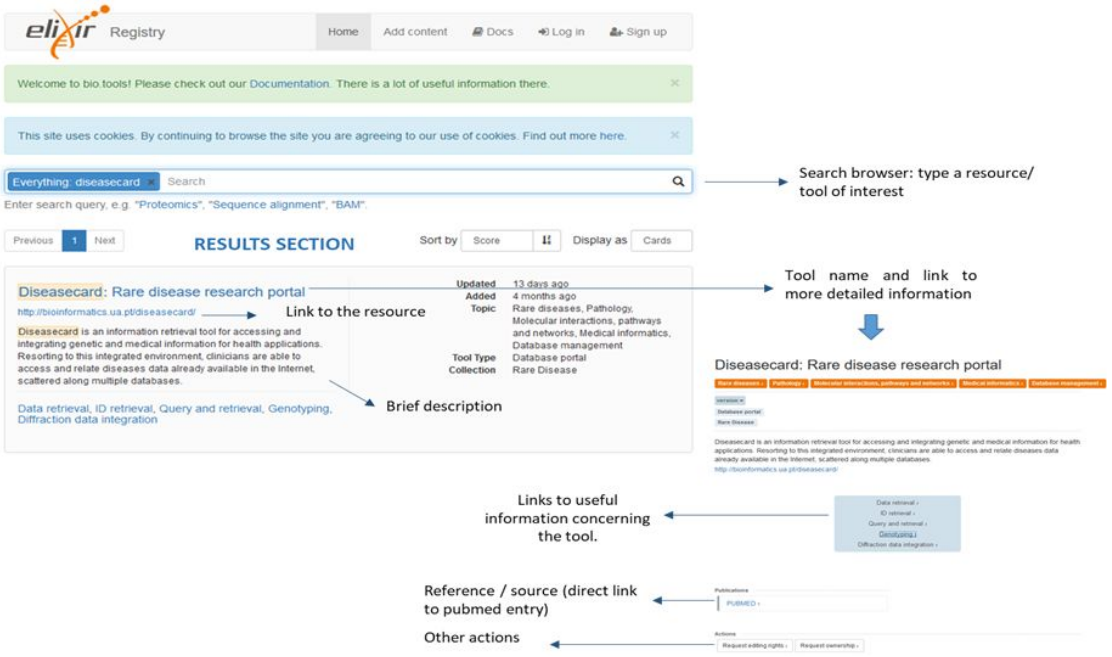


Figure 6: Use case 1: search for specific resources.

3.1.2: Search for all resources useful for a specific rare disease field research

The RD researcher / end user is interested in finding the tools that would fit better his research purposes in the RD field. The RD researcher will access bio.tools (https://bio.tools/) and type a general term such as “rare disease genomics”. The ELIXIR bio.tools application will give back a list of all the tools flagged as rare disease tools in the genomic field.

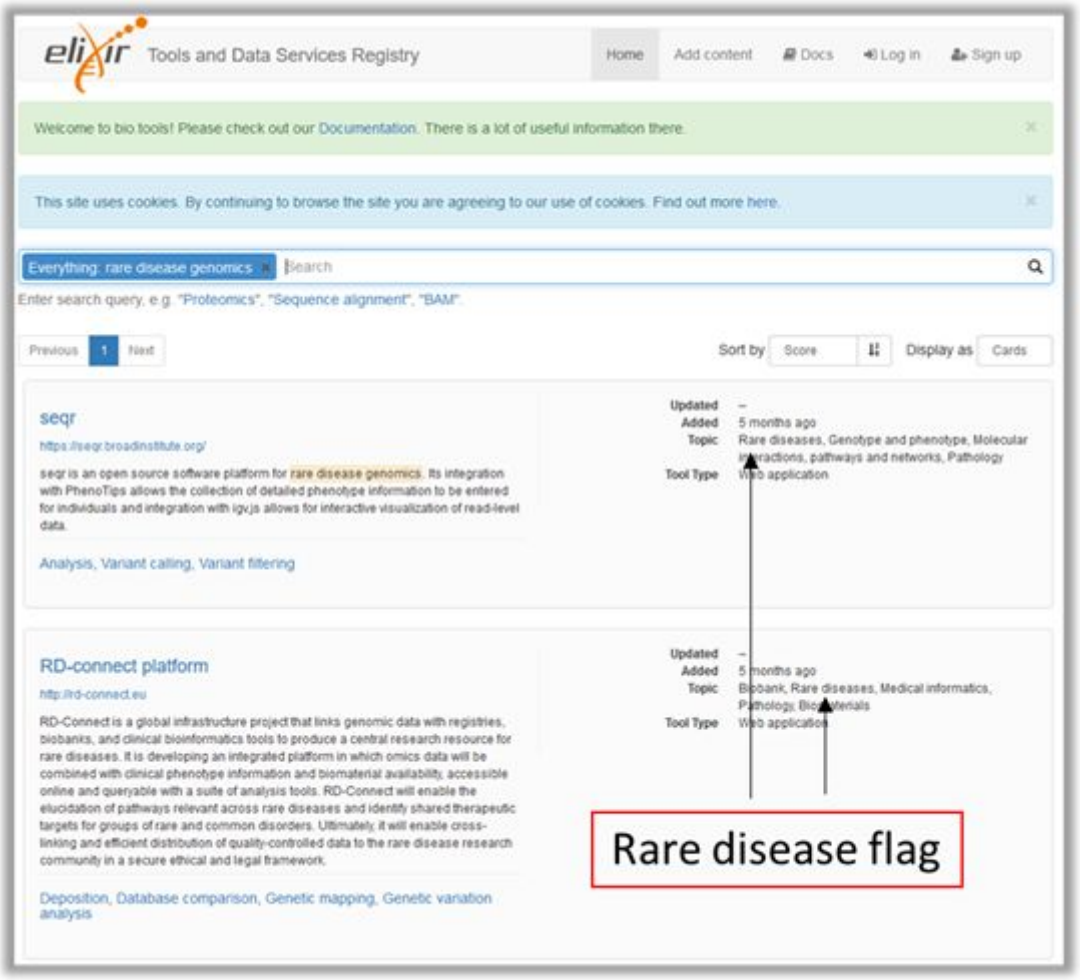


Figure 7: Use case 2: search for all resources concerning specific RD research purposes.

Moreover the researcher will be able to look at them separately and obtain a small description of the different resources, a direct link to the website and a link with additional information, references and possible other actions (Figure 7). As mentioned above, WP2 is also working on displaying benchmarking information with the associated information about the community based rating, which will be of interest for the end user to decide which tool will fit better his research purposes.

3.1.3: Search for all resources useful for rare disease research

The RD researcher / end user is interested in finding all the tools related to the RD field. The RD researcher will access bio.tools (https://bio.tools/) and select the

collection: “rare disease” (see Figure 8). The ELIXIR bio.tools application will give back a list of all the tools flagged as Rare disease tools.

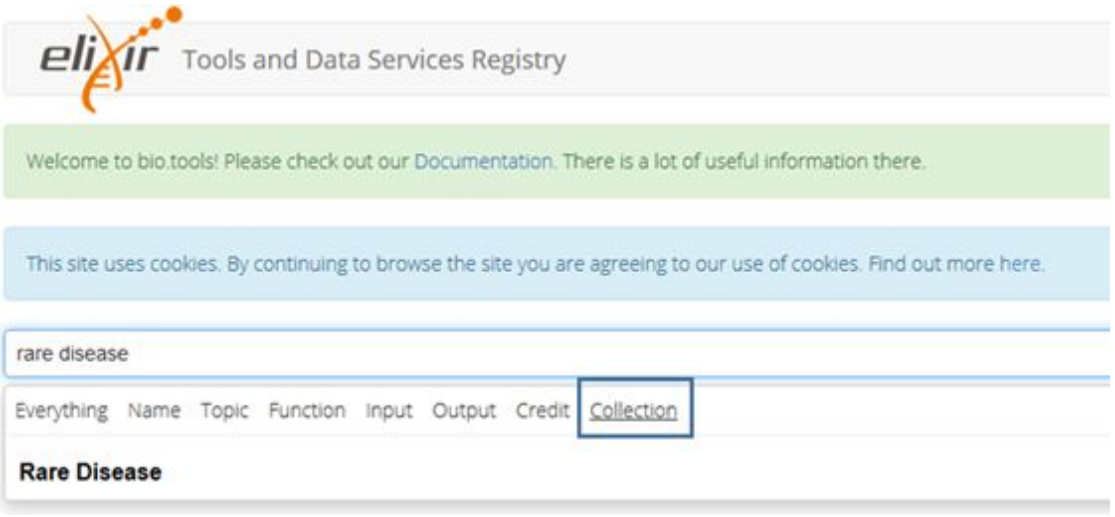


Figure 8: Use Case 3: search for all the resources useful for RD research.

3.1.4: Other examples of the usefulness of the portfolio for rare disease projects

The ELIXIR bio.tools platform is a lightweight web service with a REST interface, which provides an easy way to access the bio.tools database. They have developed an API (Application programming interface) to be used as an interface by software components to communicate with each other. For example, if within specific projects a webpage to list and describe the different tools and resources used and integrated is needed, they would be able to easily pull this information, which will also be automatically updated, through the bio.tools API.

4.- Conclusion

In the context of the ELIXIR-EXCELERATE WP8 rare disease use case, we have reviewed the ELIXIR current data resources and evaluated their usability and potential impact on the rare disease community. Through the elaboration and dissemination of a rare disease resource survey, we have started to deploy a link between the end users and the tools developers that 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 group of survey respondents was not sufficiently representative to draw general conclusions from the result. However, the individual suggestions were valuable

input in developing the portfolio of Elixir data resources and tools. Even from this limited response group, it seemed clear that compatibility and transfer of data between tools is an important issue that needs attention. The analysis and results of this survey have also been presented during the ELIXIR All Hands 2017 (21-23 March 2017 in the Barceló Aran Mantegna, Rome, Italy).

Finally we have published and made available to the rare disease community a dynamic catalogue of more than 90 resources, using the ELIXIR registry (https://bio.tools/) and in collaboration with ELIXIR- EXCELERATE WP1 and WP2. This outcome will allow users from the different countries, communities and projects to discover which are the tools available for rare disease research at a given time, with the associated information about the community based rating (see WP2), instructions for correct use and associated examples. This dynamic catalogue of resource will be constantly updated and accessible for end users through the main website or through the bio.tools API.

5.- Future plans

As we mentioned, the use of resources by the RD community is dynamic and constantly evolving, therefore, we will update and keep the survey available to ELIXIR end users and we will still be monitoring periodically respondent answers. We will also disseminate the updated survey through other channels such as the European Reference Networks (ERNs) showing the value of the current RD tools catalogue in bio.tools.

The rare disease portfolio of resources will be evaluated through the adequate selection of datasets and benchmarking strategies. This part of the process is currently ongoing in collaboration with the ELIXIR benchmarking strategy WP2 and consists in integrating these pipelines in the ELIXIR benchmarking framework to continuously monitor the selected methods with the newly generated datasets. In particular, we are currently exploring a collaboration with the Genome in a Bottle and the Benchmarking Task force of the Global Alliance for Genomics and Health (GA4GH) for benchmarking of germinal variant calling pipelines.

We acknowledge that rare disease stakeholders have multiple strategies to find tools and data resources available to them, which may be desirable to some extent but can also be confusing. We have to monitor how solutions such as bio.tools,

FAIRsharing.org, the RD-Connect portal (including ID-Cards, the sample catalogue, and Orphanet), and other platforms work together to help the RD community find tools and data resources.

Finally, we will work in collaboration with the main actors of the RD community to make them aware of the existence and value of this catalogue. These actors are intended to be 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.