RD-CONNECT Report Summary
Project ID: 305444
Funded under: FP7-HEALTH
Country: United Kingdom

Final Report Summary - RD-CONNECT (RD-CONNECT: An integrated platform connecting registries, biobanks and clinical bioinformatics for rare disease research)

Executive Summary:
RD-Connect is a global research and infrastructure resource for rare diseases (RD). Set up to overcome the siloing, fragmentation and inaccessibility of datasets from different projects, it links omics data with phenotypic data and information in registries and biobanks at both an individual-patient and whole-cohort level to enable researchers to analyse their own data and gain a complete view of their disease and patient population of interest. Data shared through RD-Connect is accessible beyond the usual institutional and national boundaries and researchers across the world can benefit from the opportunity to work with others with an interest in the same field, relate human phenotypes to a particular gene or pathway of interest, pool data to create larger cohorts, find confirmatory cases, and access samples for further study.

RD-Connect was launched on 1 November 2012 as a flagship EU project for rare diseases with six years of funding through the Seventh Framework Programme (FP7). Project partners have now successfully obtained new funding to sustain the unique resources generated by the project beyond the initial funding period, and this Final Report therefore summarises the achievements of RD-Connect as an FP7 project while at the same time pointing towards the future developments planned to keep the resources at the forefront of rare disease research infrastructure in Europe.

During the six years of FP7 funding, RD-Connect has successfully established a unique integrated platform for rare disease data and samples that brings together omics data and clinical data from participating projects with tools and services to analyse this data online. The central portal at http://platform.rd-connect.eu provides access to the suite of tools developed. This includes the RD-Connect Genome-Phenome Analysis Platform, which features a user-friendly genomic analysis interface linked to human phenotype ontology-coded phenotypic profiles for the individuals whose genomic data is accessible in the system, as well as the catalogue of biobanks and patient registries known as the RD-Connect Registry and BioBank Finder, and the RD-Connect Sample Catalogue which enables the location of rare disease biosamples. These resources are mature and increasingly used by the research community, in particular enabling clinical researchers without bioinformatics expertise to analyse their own data. Four RD-Connect-related outputs have received endorsement as “IRDiRC-Recognized Resources”: the Genome-Phenome Analysis Platform, the International Charter of Principles for sharing bio-specimens and data, the RD-Connect guidelines for the informed consent process and the FAIR Guiding Principles all originated within RD-Connect or with the participation of key RD-Connect partners.

As well as the development of a valuable and user-friendly infrastructure for rare disease research, one of the significant achievements of RD-Connect has been the way it brings together a diverse community of scientific researchers, bioinformatics experts and patients and patient representatives, all united in the task of developing new infrastructure to facilitate rare disease research and data sharing. The formal establishment in July 2018 of this community as the “RD-Connect Community” signalled the continued dedication of all participants to continue their collaborative activities and to welcome new members with the same interests. Crucially, this now includes work with the European Reference Networks for Rare Diseases: these important international initiatives are now a major feature of the RD landscape in Europe and many of them are looking to RD-Connect to provide not only the infrastructure for omics data sharing and analysis but also the expertise in data stewardship including ontologies and FAIR data principles. In addition, the ESFRI research infrastructures ELIXIR and BBMRI-ERIC are closely associated with RD-Connect. The resources developed by RD-Connect will continue to grow and develop thanks to a range of
new funding and sustainability mechanisms, in particular the Horizon 2020 project Solve-RD, the upcoming European Joint Programme for Rare Diseases, collaboration with the ESFRI research infrastructures ELIXIR and BBMRI-ERIC, and the launch of the RD-Connect Community.

Project Context and Objectives:
The rare disease context and European activities
Although individually uncommon, rare diseases (RDs) collectively affect 6–8% of the population: an estimated 30 million people across Europe. Their rarity and diversity pose specific challenges for healthcare provision and research, and for the development and marketing of treatments. Many patients with rare diseases lack timely and accurate diagnosis and even fewer receive tailored treatments that influence survival and quality of life. 80% of rare diseases have a genetic component, and the genomics revolution has brought the hope of gene-based treatments for many rare diseases a step closer. Yet bottlenecks still remain. The unmet need of the rare disease community was recognized by the European Commission, which in 2012 funded an interrelated set of three flagship projects, RD-Connect, NeurOmics, and EURenOmics, with the ambition of advancing -omics research and data sharing in line with the goals of the International Rare Disease Research Consortium (IRDiRC).

Combining and integrating genomics, transcriptomics, proteomics, metabolomics, and detailed phenotype data (phenomics) across research centres and across diseases is essential to advance knowledge of rare diseases. While competition between different research groups is a driving force to advance science, harmonisation and sharing of data is ultimately required to compare, combine and make best use of the results. This is especially true in rare diseases, where individuals with the conditions may be scattered across the world. Trans-national and trans-disease efforts are thus essential to make optimal use of resources. Patient registries, biobanks and bioinformatics analysis tools are the key infrastructure tools required for omics research. Hundreds of RD biobanks and patient registries already exist in Europe alone, and collaborative initiatives in specific disease groups (e.g. Huntington’s disease, cystic fibrosis and neuromuscular diseases) have advanced infrastructure harmonisation in several areas.

A continued bottleneck for cutting-edge research towards diagnosis and therapy development, however, is that at present these individual efforts continue to multiply while remaining largely “siloed”, with very little interoperability. Genetic information, biomaterial availability, detailed clinical information (deep phenotyping) and research/trial datasets are hardly ever systematically connected.

The objectives of RD-Connect
RD-Connect was established to tackle this siloing, fragmentation and inaccessibility of data. It aims to improve the (re)use and sharing of data and biosamples in order to facilitate research and ultimately to benefit patients living with rare diseases. As a global resource for rare disease it has developed infrastructure that links omics data with phenotypic data and information in registries and biobanks and has created new tools to analyse the data. Thanks to the user-friendliness of the tools, researchers across the world are empowered to analyse the data that they themselves submit and can benefit from the opportunity to work with others with an interest in the same field, relate human phenotypes to a particular gene or pathway of interest, pool data to create larger cohorts, find confirmatory cases, and access samples for further study.

RD-Connect began as an initiative that arose from within the rare disease research and patient community, and this meant that its objectives were those that the community itself recognised as essential to the future development of the RD research landscape. Nevertheless, to avoid duplication of effort, RD-Connect aimed to align with the activities of other large international initiatives operating in related spheres across genomics bioinformatics and data sharing. It operates under the banner of the International Rare Disease Research Consortium (IRDiRC) and collaborates closely with the major research infrastructures dealing with human data and biosamples: ELIXIR (life sciences data) and BBMRI-ERIC (biosamples and associated data). The Global Alliance for Genomics and Health (GA4GH) is another key player in the genomics environment with which RD-Connect has developed close links.

RD-Connect’s overarching objectives are to develop:
• an integrated platform to host and analyse data from RD omics research projects
• clinical bioinformatics tools for analysis and integration of molecular and clinical data to discover new disease genes, pathways and therapeutic targets in RD
• common standards and data elements for RD patient registries
common standards and a sample-level catalogue for RD biobanks
best ethical practices and recommendations for a regulatory framework for linking medical and personal data related to RD

Within the six-year work programme, these activities were further refined in a series of work packages, each with its own individual objectives contributing to the broader whole.

Integrated platform and genome-phenome analysis
The RD-Connect work programme envisaged the creation of an integrated platform for rare disease data and samples through a major international collaboration across the areas of patient registries, biobanks, bioinformatics tool development, omics data analysis, ethical, legal and social issues, and patient involvement. In particular, its core objective was to create a mechanism through which participating research projects could share and analyse data, including hosting the data from the original partner projects, EURenOmics and NeurOmics, and future IRDiRC projects. During the course of the work, in discussions with these first partner projects, it became evident that one of the most crucial aims was to lower the barriers for rare disease researchers to analyse the data they submit themselves even without bioinformatics expertise, and the development of a user-friendly interface thus grew in significance for the project. Reluctance on the part of researchers to share data due to concerns about losing ownership and being scooped in gene discovery publications was an early hurdle that RD-Connect needed to overcome. Developing a mechanism to allow the submitting clinicians and researchers to retain control over their data by empowering them to analyse it themselves thus became a key goal. In addition, going beyond genomics to explore standardisation of analysis and integration of multi-omics data, focusing mainly on transcriptomic and metabolomics use cases provided by partner projects, was a further objective. This included evaluation of RNA-Seq analysis pipelines for the identification of differential expression, variant calling and identification of fusions, in order to integrate the results with those obtained from exome sequencing.

Databases and patient registries
Genetic databases and patient registries are recognised as crucial tools for RD research. For most RD, no single institution, and in many cases no single country, has sufficient numbers of patients and resources to conduct clinical and translational research. Identifying patients with specific genotypes and phenotypes is a major constraint to patient recruitment into research projects and clinical trials. Therefore, international collaboration is essential to ascertain pathogenicity of rare genotypes, and to achieve the unified collection of RD phenotypic data, foster natural history studies and identify participants for research and clinical trials, as well as to support the safety and efficacy evaluation of potential therapies. Prior to the launch of RD-Connect it was identified that despite the large number of registries in existence, most registries faced significant challenges in terms of lack of harmonisation, lack of data sharing, lack of sustainability, lack of visibility and lack of interoperability.

The objectives of RD-Connect in the area of databases and registries were to harmonise and standardise these resources for RD by collaborating internationally to implement common registry infrastructure, standards and data elements; to enable the collection and provision of accurate, quality controlled patient data for natural history studies, identification of study participants and pharmacovigilance and the provision of well characterised and stratified patient cohorts for personalised therapies and translational research. Data in registries usually needs to remain in a federated environment for data ownership and privacy reasons, and so a major aim was to support registries in making their data FAIR (findable, accessible, interoperable and reusable) at the source.

The overall focus was thus to make high-quality clinical and phenotypic data available to the community involved in RD research. This was achieved through two parallel activities: the development of a central catalogue of registries and biobanks in order to increase the visibility of the resources, and the implementation of a data linkage plan to enable the data to be explored at source through FAIR mechanisms.

Biomaterial sharing
Access to high quality human biological materials is a prerequisite for research on rare diseases. It underpins the development of new diagnostic techniques, biomarker development, identification of potential therapeutic targets and testing therapeutic response. Biobanks play a crucial role in enabling access to samples, ensuring they are available for research and made best use of, as well as maintaining the quality of the samples from collection to distribution, and preserving the privacy and wishes of the donors.

RD biobanks tend to be specialized and small in scale; arising from the intrinsic rarity of sample collections. Nevertheless, they
serve as a key resource for RD researchers, aggregating otherwise heterogeneous and difficult-to-locate biological samples from RD patients and families. Rare disease biobanks present a rich landscape across Europe and several efforts in RD-Connect are devoted to networking activities, entailing harmonisation of quality criteria, process standardisation and compliance with ethical and legal requirements. The biobanking activities of the project were thus to harmonise and standardise RD biobanks that collect and provide standardised, quality-controlled biomaterials for translational research and integration into the RD-Connect platform, to increase access to high quality human biomaterials by developing a global catalogue of ethically collected, standardised, quality-controlled biological samples accessible to scientists engaged in research on rare diseases, and to assess the quality and standards of RD biobanks.

To achieve the objectives of the project, tasks spanned from a comprehensive mapping of existing RD biobanks, to the identification of shared datasets and standardised coding systems for biological samples, to the construction of an efficient informatics interface within the RD-Connect platform. The project aimed to streamline the processes underlying all biobanking activities into a workflow addressing each key step and its scientific, operational, ethical, and legal implications, and develop a governance model, setting priorities for access and exploitation of biobank resources.

Bioinformatics tools

High-throughput sequencing technology generates an unprecedented amount of omics data in research and diagnostic laboratories. New tools are required to make optimal use of these multi-layered and distributed datasets and interpret them in the context of the associated clinical phenotypes. The work of RD-Connect in this area had the objective of developing sophisticated bioinformatics tools to meet a range of analysis needs. The innovative systems developed were designed to be accessible either through the RD-Connect platform or as standalone software. Some tools aimed to be used by the diagnostic and research communities to more efficiently analyse NGS data (identification of disease-causing mutations in a diagnostic context and/or identification of new genes responsible for genetic diseases in a research context) or use innovative techniques such as 3D facial analysis to assist with diagnosis, while others aimed to facilitate the integration and combined analysis of multi-omics data. Reflecting the fact that genetic therapies are being developed for certain diseases, tools were required in order to select the most appropriate therapy based on the patient’s genetic background and to simplify the design of new therapeutic molecules able to induce exon-skipping, nonsense mutation read-through and trans-splicing, and to evaluate the impact of therapeutic modifications on gene expression.

Ethical, legal and social issues

Ethical, legal and social issues (ELSI) are vitally important to address in a major project like this, which involves sensitive data sharing and has the potential to change the way RD research is conducted. The ELSI-related work in RD-Connect had both practical objectives and research objectives. On a practical level, it was essential to ensure that all activities related to sensitive patient data were conducted in a secure manner and according to best ethical practices. The ELSI goals therefore included the development of a sound ethical and legal framework for the integrated platform, taking into account stakeholder views and preferences and compliant with international legal frameworks, including the relevant updates in the light of the General Data Protection Regulation. The overall need was to ensure that patient interests were protected, while recognising that patient interests include the possibility for research to be carried out on their data and thus data protection that hinders research is not in patients' best interests. The goals thus included standardised formats for information and consent procedures, sharing of data and applications to ethical review boards, together with a matrix for an expedient European regulatory framework for RD research that better protects vital patient interests and that aimed to overcome significant hurdles as witnessed by researchers involved in RD research.

In order to inform these developments, ELSI-related academic research was a further objective within the project. Research and participatory activities to assess patient preferences and expectations from infrastructure projects such as this and from research using new omics technologies in general were important aims, and this ensured an ongoing dialogue with relevant stakeholders, as well as engagement activities with patient organisations and patient groups, clinical and research networks, legislators and policymakers and pharma industry, the outcomes of which were intended to inform the framework itself. The overall objectives were thus the creation of a patient-oriented governance platform, the publication of research on patient expectations, and the development of a proposal for an expedient regulatory framework for linking of medical and personal data related to RD on a European and global level that is suitable to be used for the platform beyond the deadline of this project.
Patient involvement

Given the potential impact on patients with rare diseases of new omics technologies as well as of new mechanisms for data sharing, it is vital to include patients and patient organisations in all stages of the discussion around the research and development taking place in the project. Closely related to the ELSI activities above, a two-way dialogue between researchers and patients aimed to ensure that scientists incorporate into their work an understanding of patients' hopes and expectations for the technologies and that patients receive up-to-date information about the scope of RD-Connect. Such communication and inter-understanding should increase uptake and acceptance by providing technologies that meet patients' needs. This was made possible by activities led by EURORDIS, the alliance of RD patient organisations across Europe, in addition to the incorporation of social scientists and ethicists with proven expertise in engaging in the area of patient participation. This work aimed to ensure that all project activities remained patient-centred throughout the duration of the project by actively engaging patient representatives in relevant discussions and activities. Its objectives were to develop best practices that reflect patient interests, to engage with all relevant groups of stakeholders, including patient organisations and patient groups, clinical and research networks, legislators and policy makers as well as pharma industry, and to develop new knowledge about patient experiences and expectations in linking of medical and personal data related to RD on a European and global level.

Project Results:

Achieving an integrated platform

The RD-Connect work programme envisaged the creation of an integrated platform for rare disease data and samples through a major international collaboration across the areas of patient registries, biobanks, bioinformatics tool development, omics data analysis, ethical, legal and social issues, and patient involvement. This has been successfully achieved through the development of a suite of tools for each area accessible through an online interface. The integrated platform brings together omics data and clinical data from participating projects with tools and services to analyse this data online. The central portal available at http://platform.rd-connect.eu provides access to the genome-phenome analysis platform including the genomic analysis interface and the PhenoTips database that stores human phenotype ontology-coded phenotypic profiles for the individuals whose genomic data is accessible in the system, and the catalogue of biobanks and patient registries. The RD-Connect resources are mature and increasingly used by the research community. They have successfully secured funding that sustains them beyond the initial FP7 period. The individual results are described in more detail below.

Genome-phenome analysis platform

The RD-Connect Genome-Phenome Analysis Platform (GPAP), developed under the leadership of Ivo Gut and Sergi Beltran at CNAG-CRG, is one of the project’s flagship tools. It has developed into a uniquely powerful and user-friendly resource that enables the contributing researchers themselves to analyse the data they submit. As of October 2018, 3208 whole exomes, 363 whole genomes and 322 gene panels have been processed using the current version of the standard analysis genomics pipeline and are available for analysis in the GPAP. This number is expected to further increase by the end of 2018 since the GPAP is being used by the Solve-RD project which will be submitting 5000 datasets this year and 19,000 in total.

User-friendly data analysis

Empowering disease experts to do their own analysis online is one of the unique strengths of the system that can speed up diagnosis and gene discovery as well as provide incentives for data sharing. The GPAP is accessible to all RD researchers, even those without experience in bioinformatics. A straightforward, secure registration process allows validated researchers to access the interface to analyse and query their own data as well as data submitted by others. The data are accessible to authorized users following a predefined 6-month embargo period that gives researchers exclusive access to their own data before they are shared more widely. A researcher can select one or multiple individuals (e.g. trios or other family relationships) to study and then filter and refine the results by mode of inheritance, population frequencies, in silico pathogenicity prediction tools, gene lists and ClinVar, HPO and OMIM codes. In addition, the integrated Exomiser tool extracts HPO terms describing the symptoms of the affected patients (from PhenoTips) and prioritises candidate genes that match them most closely.

Initially launched in 2015 with data from the original partner projects NeurOmics and EURenOmics, the GPAP now welcomes any RD researcher to register as a user and submit data. A new online user registration system was launched in 2018 and the processes and documentation, including the Code of Conduct and the Adherence Agreement, were adapted to the new EU
General Data Regulation Protection (GDPR) to ensure compliance and provide greater transparency to end-users. Documentation for all the tools is available on the main RD-Connect website, and the platform homepage at https://platform.rd-connect.eu now provides links to user guides and FAQs as well as new publicly available walk-throughs and video tutorials.

The increase in user numbers has resulted in an increased demand for training, which has been provided at several levels to researchers, clinicians, patients and students through dedicated hands-on workshops and courses on genome analysis and bioinformatics. Since 2015, 17 hands-on workshops in the use of the GPAP have taken place in France, Germany, Spain, UK, Finland, Greece, Czech Republic, Slovenia, Italy and Portugal. The largest ever GPAP hands-on workshop in terms of number of attendees was organised as part of the ESHG meeting in Milan, Italy (June 2018).

Phenotypic data collection and standardisation

RD-Connect worked in close collaboration with EURenOmics, NeurOmics and other IRDiRC projects and initiatives to establish standards for phenotype descriptions to facilitate their inclusion in the central RD-Connect database and to establish mechanisms to allow the linking of phenotypic data with molecular data. This was achieved through the use of standardised ontologies and taking into account registry-related work on defining registry common data elements and implementing standardised coding systems, ontologies and classifications. After discussions with RD-Connect, both EURenOmics and NeurOmics used the Human Phenotype Ontology (HPO) to describe their patient phenotypes and worked with the HPO developers to add to the ontology any terms they needed that were missing. The decision was taken to partner with PhenoTips to provide a user-friendly mechanism for data entry using HPO. The RD-Connect instance of PhenoTips was originally set up by the PhenoTips developers in Canada and was then moved to the CNAG, where it is available as part of the genomics platform. The PhenoTips team and the CNAG prepared forms for data entry for groups of diseases according to requirements from the experts. RD-Connect is also active in other genome:phenome correlation activities including the Monarch Initiative and participated in the International Consortium for Human Phenotype Terminologies (ICHPT), which agreed on 2300 core terms that allow communication between ontologies such as HPO, PhenoDB and Elements of Morphology. Many of these phenotype-related activities are being taken to the next level through the Solve-RD project.

Data archiving

A major task for the GPAP was to establish a data workflow and to coordinate submission to archival repository. The GPAP itself stores and provides access to processed and integrated genome-phenome data, but does not aim to store raw data in the long term. Therefore, the EGA was chosen as the long-term data repository for RD-Connect. To make things easier for submitters there are two mechanisms for archival: the submitter can first submit to EGA and inform RD-Connect of the EGA ID, or can submit first to RD-Connect, which collects all the meta-data required by the EGA to allow for later submission from RD-Connect to EGA.

Standard analysis and annotation pipeline

RD-Connect mandates the submission of the raw output from the sequencing machine (FASTQ or untrimmed BAM) in order to ensure comparability of results through the use of a single standard analysis pipeline for data that may have come from multiple different sequencing providers. The incoming raw data is analysed through the RD-Connect standard analysis pipeline. A task force benchmarked several pipelines before deciding on the set of tools to use. However, the standard analysis pipeline has evolved during the project, aiming to increase sensitivity and specificity of genomic variant detection. The number of annotations provided has also increased and they are updated as necessary. Some of the annotations are provided by tools developed within the project, while also making use of leading tools available worldwide. Data releases are performed roughly on a monthly basis, and the data release is clearly indicated in the RD-Connect GPAP GUI. The latest updates to the analysis and processing pipeline have modified the way the gVCFs are merged, annotated and loaded to the system. The new pipeline is faster than before, facilitating the release of larger datasets and the addition of new annotations. Additional annotations such as gnomad and extended ClinVar have been incorporated. ELIXIR has one ongoing Implementation Study in which RD-Connect is involved as a Use Case to deploy its dockerized analysis pipeline. Another ELIXIR Implementation Study proposal has been submitted to further integrate RD-Connect and the EGA in terms of data visualisation.

GPAP technologies and user features

The RD-Connect GPAP data repository has been set up using big data technologies that index the results from the genomics standard analysis pipeline (annotated gVCFs) with Elasticsearch within a Hadoop filesystem. This allows real-time queries to
the millions of data points of the database. The results from each donor are linked to the corresponding entry in PhenoTips and can also be linked to the raw data at the EGA and the Biobank and Patient Registry in the Registry and Biobank Finder. The GPAP operates under “controlled access” for pseudonymized patient data (no identifying details). Users may analyse the data they themselves have submitted using the user-friendly interface. After performing the initial filtration and displaying the results, extensive further details are provided for each variant. Links are provided to external sources such as the ExAC, UCSC, Ensembl and NCBI browsers to view the variant in its genomic context in other populations, and to OMIM and dbSNP which provide reports on gene function and variants reported. This is continuously updated and several features have recently been added to the graphical user interface (GUI), such as MatchMaker Exchange connection to PhenomeCentral, filtering by tagged variants, analysis status, gnomad filtering, extended ClinVar filtering and filtering by pathways. Users can search by gene across all cases in the system in order to find others in which the same gene is affected – this is useful for “matchmaking” to find confirmatory cases for gene discovery and also allows researchers with an interest in a particular gene, for example basic scientists working on a particular gene in an animal model, to find corresponding human cases in which the gene is affected. A tool for establishing runs of homozygosity allows users to pick up potential consanguineous cases and focus on these regions for gene discovery. The Exomiser tool to prioritise variants based on phenotype and pathogenicity inference has been implemented. Some of the tools developed in RD-Connect such as DiseaseCard, ALFA and HSF have been integrated in the RD-Connect GPAP through a common API developed by the RD-Connect developer group. Other RD-Connect tools such as ePGA and UMD-Predictor are available to users separately, although not directly through the GPAP. RD-Connect participates in the GA4H and IRDiRC joint initiative MatchMaker Exchange (www.matchmakerexchange.org) and is part of the GA4GH Network of Beacons (https://beacon-network.org/) with both APIs being in production within the GPAP.

The GPAP has already played a critical role in the discovery of several new RD genes and phenotypes which have been published in top-level peer-reviewed journals. The collaboration with BBMRI-LPC is worth mentioning in this regard as a paradigm of good practice for European rare disease data sharing: researchers across Europe were provided with exome sequencing at no cost through a transnational access mechanism, but the project conditions mandated that biosamples must be made accessible through EuroBioBank biobanks and phenotypic data must be submitted to the RD-Connect PhenoTips instance, while the resulting sequencing data will be automatically submitted to the RD-Connect platform. This workflow not only allows the researchers themselves to analyse their own cases but also ensures that the samples and data will be accessible to other researchers in future, thus maximising the added value of the project for future research. Preliminary analysis of approximately half of these cases by the clinical genomics specialist at CNAG has resulted in molecular diagnoses of between 33-50% of cases (pending confirmation from collaborators), which is in the range expected for studies of this nature. This proves that the platform is functioning well, and those discoveries are prepared for publication. RD-Connect formally offered all applicant European Reference Networks (ERNs) the opportunity to make use of the genome-phenome analysis platform for data submission and analysis even prior to the approval of the 24 networks. The platform is open and available for use by all ERNs, both to deposit new data and to explore the existing data. The initial partner projects EURenOmics and NeurOmics contributed primarily rare neuromuscular and rare kidney disease datasets, but more recent partnerships have broadened the disease areas represented to mitochondrial, immunological, neurogenetic and other Rare Disease pathologies.

Multi-omics integration into the platform

Multi-omics analysis is an ongoing activity that was less standardised than genomics analysis at the start of the project and remains more challenging even today. The work towards an integrative omics analysis suite is described in more detail under the bioinformatics tools section below. Several task forces worked on the analysis and integration of Omics data beyond genomics. It was decided that transcriptomics would be tackled first and, afterwards, metabolomics. Transcriptomics data will be deposited at the EGA in a similar way to the Genomics data. For metabolomics, a key data resource will be Metabolights (EBI). Partners in NeurOmics and EURenOmics worked with Metabolights to upload metabolomics data and define the required data and metadata standards. It was acknowledged that it will be difficult to standardize the analysis due to the differences in the sources of data. However, researchers can use Yabi to analyze their own data. They are able to run standard analysis pipelines for any kind of experimental data. Some example workflows include (i) a typical NGS pipeline for variant annotation and filtration and (ii) a standard pipeline for metabolomics data. Yabi provides useful tools for end users and also administrator users to handle any pipelines. The first RNASeq samples were received from NeurOmics in 2018.
Data linkage and interoperability

In terms of broader data linkage and interoperability, collaborative work with other initiatives such as ELIXIR and BBMRI has been ongoing towards enabling data linkage with biobank sample data and registry phenotypic data. Work on data interoperability principles has been taken forward through the Linked Data and Ontologies Task force led by LUMC described in more detail under the databases and patient registries section. In this regard we developed close links with the ELIXIR Excelerate interoperability experts and worked with the FAIR data publishing group to further refine the standards for FAIR (Findable, Accessible, Interoperable, Reusable) data publication. This is important as several key resources external to RD-Connect such as the gene:disease associations, the phenotype:disease associations, DisGeNet and the Human Protein Atlas are available in a FAIR nanopublication format (see https://datahub.io/organization/nanopublications and http://rdf.biosemantics.org/).

In a joint activity between the platform, ELSI and biobanking partners and in collaboration with the joint task force on privacy-preserving record linkage established by IRDiRC and the Global Alliance for Genomics and Health (GA4GH), RD-Connect explored potential mechanisms to link records on the same research participant in different databases without revealing the individual’s identity (PMID: 30059313). With the data from the BBMRI-LPC project, RD-Connect started to pilot the use of EUPID (a Privacy Preserving Record Linkage solution) to anonymously connect information from the same patient available in different resources (e.g. sample in a given biobank and genomic data in the RD-Connect GPAP).

The RD-Connect platform will continue to provide services beyond the RD-Connect funding period thanks to projects that will use it as their main data collation and analysis resource (e.g. Solve-RD) or that will build on it to provide further functionalities and integration (e.g. EJP-RD). Additionally, the Titinopathies consortium is planning to deposit around 5000 datasets. IRDiRC has accepted the RD-Connect GPAP as a recognized resource and the E-Rare JTC2018 and German BMBF RD calls recommended or encouraged the use of the RD-Connect GPAP to applicants submitting responses to these calls. Continued collaboration between RD-Connect, ELIXIR and BBMRI-ERIC has also embedded the RD-Connect GPAP into European rare disease data infrastructure. To date, the GPAP has helped to reach diagnosis for at least 340 cases, and functional studies are being carried out for over 45 additional cases. RD-Connect has facilitated data sharing and analysis enabling the discovery of over 120 new disease-causing genes by NeurOmics and EURenOmics (PMID: 29487416). Furthermore, 10 new candidate genes from the BBMRI-LPC call are currently being evaluated, and more candidates may arise as the analysis continues.

Databases and patient registries

Patient registry activities were led by Domenica Taruscio (ISS, Rome). These activities focused on the development of the Registry and Biobank Finder, as well as on FAIRification of data, integration, quality assessment and interoperability.

Registry & Biobank Finder

One of the main achievements of the project in this area was the development of the Registry & Biobank Finder, a unique catalogue of registries and biobanks that increases the visibility of those important resources and provides a mechanism for researchers, particularly those from European Reference Networks, to locate the data and samples they require. The online catalogue was developed using the open source Liferay® portal system in a collaboration between ISS and developers at MUG. The general architecture of the catalogue was drafted during a meeting in December 2013 between ISS, MUG and Telethon. The starting point for the data collection was the inventory of existing registries/databases created at the beginning of the project, which gathered information from different sources (ORPHANET, EPIRARE, HQIP, EURORDIS, Treat NMD, CORDIS, E-RARE, BBMRI, EuroBioBank, Telethon Network of Genetic Biobanks). This inventory was imported into the Liferay software as a list searchable by disease name, country, type of database (biobank or registry) ORPHAcode, ICD10 number and OMIM code, synonyms and other keywords (e.g. words in the database name). Biobanks and registries involved in NeurOmics and EURenOmics and other biobanks/registries in the unified list were invited to create an ID card for their resource and make their content visible to the community of RD researchers. The Finder now includes over 360 registries and biobanks and has been used by researchers and ERNs to identify relevant resources. The preliminary focus of the work to include resources focused on those participating in NeurOmics and EURenOmics, and subsequently those involved in ERNs. This also prompted an update to the system in order to be able to

Registry quality standards

A further important achievement was the development of recommendations for registry quality. This work involved gathering a group of subject matter experts, including members of patient organizations, with wide experience in the field of RD
registries to form an expert working group to review existing guidelines. The objective of the expert working group was to make a list of recommendations to be used as a framework for improving the quality of RD registries. After reviewing several guidelines on RD registries, the group came up with factors that influence the overall registry quality and a list of recommendations was elaborated for each aspect. The recommendations made by the experts were focused on: establishment of the registry governance, identification of the right data sources, specification of Data Elements, Case Report Form, Standardisations, construction of IT infrastructure complying with FAIR principles, production of data quality and to the dissemination of a quality information. Other topics such as developing adequate documentation, training of staff and providing data quality audit are considered also essential for improving registry quality. A publication (Kodra et al, PMID: 30081484) provides the recommendations in full.

Data linkage plan
Interoperability of datasets stored in different repositories is essential to enable data sharing between registries and integration of registry data with other data types. To achieve this goal, RD-Connect has drawn extensively on the expertise of the data linkage experts in the Leiden University Medical Centre (LUMC) and University Medical Centre Groningen (UMCG) teams, who developed a data linkage plan for the project as a whole, with a particular focus on cross-resource data linkage including interoperability for patient registries. This includes mechanisms for implementing FAIR data principles to make the datasets stored in patient registries Findable, Accessible, Interoperable and Reusable.

Aiding the rare disease community in developing robust infrastructure that supports data integration needs to take into account that there are over 6000 rare diseases with multiple types of resources (biobanks, registries, omics) across countries. Researchers should be able to combine data from these resources, because of the relative sparsity of the data. Maintaining a centralized warehouse at this scale, and with this kind of sensitive data, is neither feasible nor ethically or legally acceptable. We need to provide solutions that can scale-up to be adopted by thousands of resources ‘at the source’ and facilitate cross-resource analytics at the level of the data itself. This mitigates the high costs of researchers spending too much time reconciling data ambiguities, while previous reconciliation efforts cannot be reused. Working in close collaboration with data stewardship and interoperability experts from ELIXIR, a ‘rare disease data linkage plan’ was written and endorsed by stakeholders in the rare disease community and infrastructure experts. The signatories to the plan committed to making rare disease resources findable, accessible, interoperable, and reusable by humans and computers (FAIR) at the source. It was supported by RD-Connect, Elixir, CORBEL, BBMRI, FAIRDict, and ODEX4All, and rare disease patient organisations have shown strong willingness to co-invest. The plan provides recommendations for data annotation and exchange, and tooling. Early design decisions include using the FAIR data API, ontologies, and linked data to enable cross-resource analysis. The project incrementally provides ontology recommendations, starting with the HPO for phenotypes and ORDO for diseases. This work has been integrated into several conferences, workshops, courses and tutorials for patient registries and is continuing under the auspices of the EJP-RD, where FAIR principles for data stewardship play a leading role. Within RD-Connect, the International Summer School for Rare Disease and Orphan Drug Registries and the ‘Bring Your Own Data’ workshops significantly influenced the development of the FAIR Guiding Principles as applied to rare disease. This also resulted in the achievement that in November 2017 the FAIR Guiding Principles received the label IRDiRC Recognized Resource.

Biomaterial sharing
Biobanking activities were led by Lucia Monaco (FTELE, Milan). These activities focused on the development of the Registry and Biobank Finder and Sample Catalogue, as well as on FAIRification of data, integration, quality assessment and interoperability.

In the area of biobanking, the main activities centred on tool improvements and data loading, as well as training and engagement. The RD-Connect Sample Catalogue was placed in production on the https://samples.rd-connect.eu domain. Datasets were received from a number of EuroBioBank biobanks and mapping is underway, with a substantial number of samples in the staging area, ready to be published. The first release of the Sample Catalogue had more than 7000 samples listed, covering 90 rare diseases. With the release of the Sample Catalogue, training webinars were arranged for biobank owners on data preparation. The procedures and invitations for incoming biobanks to use RD-Connect were part of the publication “The RD-Connect Registry & Biobank Finder: a tool for sharing aggregated data and metadata among rare disease researchers” (Gainotti et al., Eur J Hum Genet, May 2018, DOI:10.1038/s41431-017-0085-z).

Rare Disease biobank community
RD biobanks have particular need to network and link with other existing entities. There is already a solid network of RD biobanks - EuroBioBank - driving excellent biobanking practices since 2003. EuroBioBank is an inclusive network of RD biobanks with a long experience in harmonisation, networking, and driving biobanking excellence together. It contributed to shaping research infrastructure development via RD-Connect by expression of needs, sharing workflows and piloting tool implementations. EuroBioBank members also contributed to the RD-Connect Panel for Biobank Assessment. In 2016, EuroBioBank agreed to be the de facto biobank network of RD-Connect. As a part of the integration and rebranding to RD-Connect EuroBioBank, the EuroBioBank website (www.eurobiobank.org) was revamped and relaunched in October 2017. The new EuroBioBank website has many direct links to the RD-Connect tools (documents, Sample Catalogue, Registry & Biobank Finder), and emphasises its position as the RD-Connect biobank network. During the timeframe of RD-Connect, EuroBioBank grew from 16 to 25 biobank members. It currently holds the only known major collections of biological samples from RD patients and families in Europe, with more than 150,000 biological samples covering more than 850 diseases in the catalogue (data 2016).

RD-Connect Sample Catalogue
A centralised catalogue of RD samples listing collections from multiple biobanks helps to alleviate the difficulties in locating rare samples. The flagship tool allowing researchers to locate RD samples, RD-Connect Sample Catalogue samples.rd-connect.eu was successfully created. The catalogue was built using the latest data management guides, standards and technologies, and is a unique tool in providing sample-level information. The first release of the Sample Catalogue had more than 7000 samples listed, covering 90 rare diseases. Complementary training webinars on data preparation for publication were arranged for biobank owners with the release of the Sample Catalogue. Besides biological samples, users can also find the information about the hosting biobanks easily via links to the RD-Connect Registry and Biobank Finder.

RD-Connect Registry and Biobank Finder
As described in the registries section above, the RD-Connect Registry & Biobank Finder(http://catalogue.rd-connect.eu) is a tool listing registries and biobanks that are collecting data on patients affected by a RD (Gainotti et al., Eur J Hum Genet, May 2018, DOI: 10.1038/s41431-017-0085-z). It is an enriched directory, containing metadata such as the number of cases or biological samples included in each database and documentation sections on case report forms, templates or access documents. Researchers can search the directory for biobanks or registries jointly or separately, by biobank/registry name, disease name, ORPHA code, OMIM and ICD-10 codes, synonyms, and other keywords related to the database. The Registry & Biobank Finder was built using the Open Source Liferay Portal Framework. Aside from being a portal for database owners to make their resources known, it provides information on the existence of specific samples and location coverage of the data sources. An important feature of the Finder is the direct management of the information by the resource owners. RD registry and biobank managers are provided with personal user accounts with which they can login and keep the information on their resources and contacts up-to-date and correct.

Quality Biobanks & Workflow for inclusion
The quality assurance of biological samples and biobanks are paramount to ensure meaningful research results and reproducibility. RD-Connect is encouraging RD biobanks to make their sample collections available to the scientific community through the incorporation of their sample collections into the RD-Connect Sample Catalogue. However, to provide confidence to users that the biological samples were collected with compliance to best practices and requirements, interested biobanks have to declare their adherence to minimal conditions, the adoption of standardisation and harmonisation measures and the adherence to general ELSI principles. To this end, a dedicated procedure entailing specific assessment criteria and review of candidate biobanks was emplaced. The evaluation process is managed entirely via the Registry & Biobank Finder and contains 4 stages: online registration, application form, evaluation and feedback. The applications are reviewed by a panel of experts dedicated to the assessment of new biobanks (RD-Connect Panel for Biobank Assessment). The biobanks are informed of the final decision by electronic means and receive an Assessment Report. The Assessment report contains valuable suggestions and advice for candidate biobanks on how to improve areas of operations. If a biobank is successfully approved by the panel it will be accepted in the Finder, appear in the Directory and finally, load its sample collection data into the Sample Catalogue.

Dissemination to -omics projects

Page 10 of 39
Research and Innovation
RD-Connect EuroBioBank partnered with CNAG - CRG in the 2016 BBMRI – LPC Whole Exome Sequencing (WES) Call. In particular, the Call offered a unique opportunity to genetically diagnose rare disease patients with DNA samples deposited in the EBB network. The Call provided free-of-charge WES and bioinformatics analysis for a total of nearly 900 samples, including patients and their relatives. For the 17 awarded projects, RD-Connect partners and EuroBioBank provided DNA as well as biobanking services for those samples that were not stored in biobanks. The partnerships between researcher, biobanks, sequencing services, bioinformaticians, analyses and data sharing platform via this Call was a successful demonstration of a working RD-Connect ecosystem for rare disease research.

Collaboration with patient organisations

Associate Partner TNGB authored a flagship publication in October 2016 on the alliance between genetic biobanks and patient organisations (Baldo et al., 2016; DOI: 10.1186/s13023-016-0527-7). TNGB proposed and worked with a model on how RD biobanks and patient organisations can formalise collaborative agreements to promote awareness in biobanking services, potentials of biological samples and setting up of disease specific sample collections. Currently there are 14 patient organisations actively collaborating with TNGB under the published agreement model. Overtime, more than 2000 biological samples were deposited under these formal agreements (data March 2018).

Sustainable tools and operations

Both the Sample Catalogue and the Finder were built with technical and data interoperability in mind, so that these tools can be compatible to other existing platforms and better long term sustainability. The adoption BBMRI-ERIC Negotiator as the Sample Catalogue workflow management tool is an example of such interoperability. In the last year of the project, sustainability discussion on RD biobank infrastructure and EuroBioBank with BBMRI-ERIC led to the preparation of a “Vision Paper”, which described the state-of-the-art on RD-Connect biobank tools, network, activities and a recommendation on the resources required to maintain them. These discussions led to the agreement that BBMRI-ERIC will maintain the operations of the Sample Catalogue and Finder tools. Rare Disease continues to be a part of the BBMRI-ERIC work program, in which the special case is made to allow RD biobanks to join from non-member countries.

Bioinformatics tools

In addition to the central resources offered through the platform interface, RD-Connect partners have under the leadership of Christophe Beroud (AMU) developed a number of bioinformatics tools to assist researchers in omics analysis and therapeutic target identification. The goals of the bioinformatics tools activities within RD-Connect were to develop various suites of tools to extract knowledge from high throughput experiments, clinical registries and biobanks. Twenty-one deliverables were successfully completed and allowed the creation of highly innovative systems accessible either through the RD-Connect platform or as standalone software. They can be used by the diagnostic and research communities to more efficiently analyse NGS data (identification of disease-causing mutations in a diagnostic context and/or identification of new genes responsible for genetic diseases in a research context, e.g. PMID 29860484) and to facilitate the integration of multi-omics data (e.g. PMID 29929540). In addition, tools were developed to simplify the design of new therapeutic molecules able to induce trans-splicing, and to evaluate the impact of therapeutic modifications on gene expression.

Development of a DNA variant analysis and prioritisation suite

The identification of mutations responsible for rare genetic diseases has been facilitated by the development of the Next Generation Sequencing (NGS) technologies. We have thus moved from a gene by gene approach to gene panels, whole exome sequencing (WES) and now whole genome sequencing (WGS) approaches. Concomitantly, we observed a switch in analysis limitations from the technical ability to produce enough sequences to the "data deluge". In fact, when we compare each WES dataset to the reference genome, we observe an average of 50 to 100,000 mutations, this rising to 3 to 4 million for each WGS. Here we created DNA variant prioritisation systems to efficiently annotate and filter mutations to rapidly pinpoint a handful of candidate disease-causing mutations. These systems were mainly designed for mutations located in exons and introns of known genes and therefore useful for gene panels, WES and WGS. In parallel, because mutations can also be localized in 5' and 3' untranslated region (UTR) that may regulate genes expression, another system was specifically developed to handle such mutations. RD-Connect partners produced, evaluated and released fully functional innovative tools to solve this problem. These tools are:

• The UMD-Predictor system (http://umd-predictor.eu)
• The Human Splicing Finder system (HSF) (http://www.umd.be/HSF3/)
These systems have been continuously improved and fully were described elsewhere. In summary these tools have the following characteristics:

**UMD-Predictor**: this system was designed to predict the pathogenicity of exonic substitutions, which result either in nonsense, non-synonymous or synonymous changes at the protein level. It is based on an innovative combinatorial approach that aggregates data related to splicing signals, evolutionary data, biochemical substitutions matrices, mutant position within the protein and allelic frequency in the general population. It provides with predictions ranging from 0 to 100, divided into four classes: 1) <50 polymorphism; 2) 50 to 64 probable polymorphism; 3) 65 to 74 probably pathogenic mutation and 4) >74 pathogenic mutation. All evaluations demonstrated that this system is the most accurate, fast and reliable system to handle NGS data. It has been published in 2016 (Salgado, D. et al. Hum. Mutat. 37, 439–446 (2016)) and was well cited.

**Human Splicing Finder**: this system was designed as a one-stop-shop for splicing signals. It therefore combines multiple algorithms and matrices to identify splicing signals (5' splice site (5'ss) or acceptor splice site, 3' splice site (3'ss) or donor signal, the branch point as well as auxiliary sequences known to either enhance or repress splicing: Exonic Splicing Enhancers (ESE) and Exonic Splicing Silencers (ESS)) and predicts the impact of mutations on these signals. Because an exhaustive analysis of the impact of mutations on splicing signals usually results in large amount of data making interpretation difficult, it contains an expert system able to handle this flow of information and produce an easy-to-understand prediction. Various evaluations have been performed and demonstrated that HSF is very efficient to predict the impact of mutations on splicing signals (98.8% accuracy), branch points (100% accuracy) and auxiliary signals (85.2% accuracy, 92% specificity and 79% sensitivity). With the availability of the expert system to facilitate data interpretation, HSF rapidly became the reference system for the prediction of mutations impact on splicing signals.

**Variant Annotation and Filtration Tool**: this stand-alone application written in Java can be used on most computers (various binaries are available to download for Mac, Windows and Linux operating systems). It provides a full graphical interface and includes unique features to improve mutation annotation and prioritization. It combines classical data (phylogenetic, conservation and protein structures) with additional information at variant, gene and phenotype levels. In addition, it is one of the few systems able to combine small (single nucleotide variations, small insertion/deletions) and large rearrangements (copy number variations) to get a comprehensive picture of the individual genome. With VarAFT, users can easily annotate, filter and perform breadth and depth of coverage analysis from their data without computer programming skills and with limited hardware requirements, to efficiently identify disease-causing mutations as demonstrated in various situations. It has been published in 2018 (Desvignes, J.-P. et al. Nucleic Acids Res. 46, W545–W553 (2018)) in the Nucleic Acid Research Web Server Special issue and is already internationally widely used.

**ALamut Functional Annotation**: This tool aims to identify genetic variations located in non-coding DNA regions that may be involved in various gene regulation processes. It allows users to explore a large number of functional elements in noncoding regulatory regions of the human genome providing opportunities to identify and investigate noncoding variants identified in NGS experiments. It contains information related to transcriptional and post-transcriptional mechanisms such as promoter regions, transcription factor binding sites, enhancer regions, CTCF binding sites or CCCTC binding sites (insulators), microRNA target sites and chromatin state landmarks and epigenetic modifications. A web user interface was designed to access the ALFA database and an API was released to directly connect the ALFA tool to the RD-Connect Genome-Phenome Analysis platform. Several machine learning methods (CADD, GWAVA, DANN, FATHMM-MKL) have been recently developed to evaluate arbitrary genomic single-nucleotide variants with respect to their potential to affect gene regulation. These methods compute their own scoring system based on numerous genomic features (from 20 to more than 100 features) to prioritize non-coding variants. All of them use genomic annotations from the ENCODE project available from UCSC.

As regulatory regions are so diverse, ALFA uses a different approach by defining rules and constructing region-based strategies to prioritize non-coding variations. The framework focuses on a subset of well-characterized regulatory regions gathered from the Ensembl regulatory project: promoters, TFBS and microRNA binding sites. To evaluate and validate this strategy, IBS will have to build a dataset of published disease-causing or functional regulatory variants on these regulatory regions. These works are in progress.

These systems have been integrated into the RD-Connect Genome-Phenome Analysis platform (UMD-Predictor, HSF and ALFA).
or were instrumental to its creation (VarAFT). Because of the high quality of the systems, a spin-off has been created to ensure the valorisation of the UMD-Predictor and HSF systems. Concomitantly, IBS will ensure the valorisation of the ALFA system through its own product line.

Development of integrative omics analysis suite
Multi-omics pathway analysis is a reality since more and more data is being generated at different levels in an experimental setting. For example, previously when blood samples were collected from patients only gene expression would be measured, but that is now being complemented with for example metabolomics and lipidomics analysis of the same samples. Therefore, creation of tools and workflows for mining and integration of genomic, transcriptomic, proteomic and metabolomic data are being actively developed by the scientific community. The goals of this task were to develop innovative systems to handle multi-omics data, validate these tools through their evaluation using various use-cases.

Unambiguous pathway analysis is dependent on correct mapping of genes, transcripts, proteins and metabolites to database identifiers. While the bioinformatics community has devoted a lot of effort to successfully standardize gene and protein names and identifiers, the metabolomics community is still struggling. Together the RD-Connect partners produced, evaluated and released fully functional innovative tools to solve this problem. These systems have been continuously improved and fully described elsewhere. in summary the main achievements are:

Workflows for pathway annotation: They have been designed using the Taverna workbench version 2.5 and the WikiPathways Web Service API, for pathway annotation of metabolomics and transcriptomics data. The efficiency was demonstrated with a Huntington's disease patients use case. These workflows accept ChEBI identifiers or ENSEMBL identifiers. All workflows were annotated according to the Best Practices for Workflow design.

Workflows for pathway overrepresentation analysis of metabolomics and transcriptomics data: Five workflows using Taverna and WikiPathways as above, and an R script using the dhyper function from the Hypergeometric R package (version R 3.1.2 "Pumpkin Helmet"). These workflows retrieve the information that is necessary to calculate the hypergeometric probability that a specific pathway is significant given a list of significant ChEBI metabolite identifiers or ENSEMBL gene identifiers. The workflows and R scripts are all available online.

Literature Wide Association Study” (LWAS): The LWAS was based on the associative information contained in 17 million MEDLINE abstracts. Of the 417,561,711 possible gene-disease pairs (19,113 genes x 21,847 diseases in our thesaurus) more than half (213,489,335) lacked sufficient literature representation to build a concept profile for either one or both of the concepts in the pair. Match scores were thus computed for the remaining 204,072,376 pairs and normalized to a percentile. Apart from explicit associations of genes with diseases, as reported in databases or in the literature, there are also many indirect associations that can be inferred from the literature: They may have the following type: when disease A is associated with gene B and gene B is associated with gene C, disease A may also be associated with gene C. We refer to the AC relations as implicit associations. Explicit associations account for only 0.73% of the total (1,479,895 co-occurring gene-disease pairs) but have the highest match scores (association strength). The distribution of the explicit associations peaks in the 99.4 percentile, the vast majority of the associations, however, are implicit. The RD-Connect partners have demonstrated the application of LWAS in rationalizing the implicit links between genes and diseases/phenotypes using high-ranking gene-disease/phenotype pairs. Together, through various use cases, they demonstrated how LWAS can help interpret the biomedical significance of new gene candidates causing disease or contributing to pathophysiology. The relative proportion of Type IV associations is likely to decrease with lower percentile scores. However, as the vast majority of gene-disease/phenotype associations have yet to be explicitly stated the implicitome provides the broadest possible knowledge base in which gene-disease/phenotype pairs can be interpreted. The current set of gene-disease and gene-phenotype associations will be integrated with the RD-Connect platform. Since efforts to develop a Common API in the RD-Connect platform design are already underway, we plan to adopt this API as soon as the first API specification is available. Our gene-disease and gene-phenotype datasets will then serve as a pilot to test the Common API in a practical, real-world use-case. Guidelines for best practices and web applications and services for semantic enrichment and semantic data integration of genotype-phenotype data: RD-Connect partners released guidelines for FAIRification of RD data sources through a general FAIRification process. As guidelines should be widely adopted by the community, this initiative was developed in close collaboration with the ELIXIR network. In addition to the general principle, it is important to consider that RD genotype-phenotype data are sparse and highly distributed. Therefore, interoperability is crucial. This led to the development of an RDF
Scaleus is a semantic web tool focused on the semantic integration and enrichment of data developed by the UVAR partner. Its main features are: i) storage, ii) data integration, and iii) inference. Scaleus is powered by a transactional database (TDB) for RDF storage. This provides a high performance and transactional triplestore capable of handling efficient queries by means of a native store. It was used to create an RD application “The demonstrator” (http://bioinformatics.ua.pt/rd-connect-demo/) in the context of the RD-Connect project through an ELIXIR (www.elixir-europe.org) implementation study. The demonstrator application enables queries across RD resources related to the “Ring14 syndrome”, a very rare chromosomal abnormality, following the FAIR data principles. Workflow for semantic integration of multi-level -omics data: The objective was to develop a workflow for integration of tools and web services based on new extensions of Semantic Web technology to provide intuitive and real-time access to data and databases deemed relevant based on meta-data descriptions on patient phenotypes, experimental design and analysis methods. The workflow was designed as a novel interoperability architecture following the FAIR principles that combines three pre-existing Web technologies to enhance the discovery, integration, and reuse of data in repositories. The infrastructure consists of three main building blocks: 1) a FAIRifier tool; 2) a FAIR Data Point to provide high level metadata descriptors about data deposits, and to provide instructions to access various distributions of data sets; and 3) a FAIR Data Search Engine based on harvesting metadata from FAIR Data Points. It was designed as a 5-layered metadata schema comprising repository metadata, i.e. record containing information about the FAIR Data as a data repository; catalogue metadata, i.e. record containing information about the data catalogue(s), i.e. the collections of datasets; dataset metadata, i.e. record containing information about each individual dataset; distribution metadata, i.e. record containing information about each dataset’s distributions; data record metadata, i.e. record containing information about the dataset’s record, i.e. the internal structure, types, their relations (the semantic model). The FAIR Data Search Engine [https://github.com/DTL-FAIRData/FAIRSearchEngine] harvests the metadata available on FAIR Data Points or compatible data repositories, indexes them, and provides a search interface.

Development of integrative -omics Analysis Suite: Development of integrated approaches for data analysis and interpretation linking transcript, protein and metabolite bio-signatures to clinical phenotypes was desperately needed. The RD-Connect partners thus developed new approaches to construct multi-omics bio-signatures for objective monitoring of patient condition and response to treatment depending on the amount and type of data available and tested them on transcriptomics, metabolomics, lipidomics and proteomics data from mice and human. A web tool to perform the integration of these different types of omics data is now available for local installation at https://github.com/LUMC-BioSemantics/crosslinkWGCNA. After performing the standard WGCNA steps, user uploaded projects can be integrated by correlating the module eigengenes pairwise. The resulting correlation values are visualized in a circle layout. Additionally, modules are hierarchically clustered, and the resulting trees are visualized alongside the modules to aid interpretation.

The WGCNA integrative omics web tool allows for unsupervised analysis of multi-omics data. Groups of co-regulated molecules from different omics can be correlated, related to phenotype information, and further investigated for disease relevance using physiological processes annotation. One drawback is that WGCNA demands at least 15 samples for a reliable analysis, so for experiments with lower number of samples this method is not an option. For such experiments differential expression analysis followed by pathway overlap analysis is an option. Analysis of three different multi-omics datasets, one on mdx mice from the Neuromics project, one on beta-thalassemia from the BRFAA associated partner, and one on Huntington’s disease from the LUMC revealed inter-omics correlations that should be investigated further and followed up with smaller validation wet lab experiments.

In a separate activity, Australian RD-Connect partners developed the Yabi workflow environment. The Yabi web-based analytic environment allows the design and deployment of a flexible and generic bioinformatics framework for specialized data analyses thanks to a modular design and architecture. It is scalable, highly configurable, user-friendly, secure and open source, which facilitates seamless and transparent access to heterogeneous supercomputing and cloud environments. Yabi provides an analysis workflow environment that can create and reuse workflows as well as manage large amounts of both raw and processed data in a secure and flexible way across geographically distributed computing resources. These characteristics make Yabi an attractive system to globally coordinate bioinformatics efforts in RD. Yabi workflows were developed to handle genomic data as well as metabolomics data. Metabolomics has emerged as an important functional genomics tool that can...
significantly contribute to the understanding of complex metabolic processes. A number of analytical technologies have been utilized and successfully employed to analyze metabolites in many different organisms, tissues, and fluid. Some of these techniques include GC-MS (Gas Chromatography Mass Spectrometry), LC-MS (liquid Chromatography Mass Spectrometry), NMR (Nucleic Magnetic Resonance), CE-MS (Capillary Electrophoresis), EI-MS (Electrospray Ionization Liquid Chromatography) and several combinations of technologies such GC x GC-MS, or tandem MS. Since no single analytic technique covers the entire spectrum of the human metabolome, it is becoming a practice to use more than one platform for metabolomics studies.

Two particular tools have been added in Yabi for processing GC-MS and LC-MS data for non-targeted analysis. PyMS that is a Python toolkit for processing GC-MS data and XCMS for LC-MS data processing. Reports from both of these tools can be used for further statistical analysis and data integration.

Development of a clinico-genomic knowledge discovery suite

When next-generation re-sequencing-based PGx (pharmacogenetics and pharmacogenomics) testing becomes widely available, it will require a substantial effort to translate this genomic information into clinically meaningful guidelines. In real-life situations, the PGx clinical scenarios are truly complex, which often pose significant dilemmas to the medical professionals regarding the selection of a treatment modality. Within RD-Connect, the aim was to support drug-prescribing decision makers and provide a practical solution in incorporating PGx knowledge into routine clinical practice. The solution developed within RD-Connect was an integrated information system, to serve as an electronic PGx assistant - ePGA.

The scope of ePGA is twofold: first, to facilitate and enhance identification and evidence-based documentation of (existing or newly discovered) PGx gene-drug-phenotype associations and second, to translate and transfer well-documented PGx knowledge to clinical implementation aiding to the rationalization and individualization of therapy. The ePGA is a ‘one stop shop' Web-based platform to ease the processing, assimilation and sharing of PGx knowledge, and facilitate the aggregation of different PGx stakeholders’ perspectives. The platform takes advantage of interoperable and flexible bioinformatics and advanced information processing components that are able to serve two major PGx tasks: (a) to offer personalized diagnostics based on reliable genomic/genetic evidence, and (b) to reduce healthcare costs by increasing drug efficacy and minimizing adverse drug reactions.

New therapy feasibility studies

The research landscape in the RD field is the archetype of translational research ranging from gene discovery to new therapeutics developments to ultimately cure patients. This has been clearly emphasized in the IRDiRC goals. With the emergence of new genotype-based therapeutic approaches, it became clear that bioinformatics systems could play a major role by linking concepts to big data. The task 4.4 was originally designed to develop such bioinformatics systems. It led to the development of 3 tools:

- The SKIP-e system (http://skip-e.geneticsandbioinformatics.eu)
- The NR-Analyzer
- The Crawfish system (http://crawfish.geneticsandbioinformatics.eu)

These systems have been continuously improved. In summary these tools have the following characteristics:

SKIP-e: this system was designed to assist researchers in selecting the best region of interest to design Antisense Oligonucleotide (AON) sequences to induce exon skipping or more globally alter exon recognition. AON are short, single stranded DNA or RNA molecules that can bind to a specific RNA region of interest leading to the inactivation or the degradation of a specific transcript. These molecules are either used as therapeutic agents to restore the production of a functional deleted protein e.g. DMD or as tools to study gene function through the inactivation of a protein through Nonsense Mediated Decay (NMD). AON are small molecules that can easily be produced and therefore potential therapeutic molecules of high interest. So far, their selection has mostly been empirical and SKIP-e was designed to rationalize their selection and allow high throughput analysis. SKIP-e now contains all potential AONs of sizes ranging from 15 to 40 nt that can target any exon of any human transcript. The 33.3 billion (33,290,445,688) of AONs were thus created and pre-computed to search for specificity (to avoid unexpected off-target effects) and efficacy (exon recognition is altered only if splicing signals are hidden by the AON). In parallel an AON database was created to collect all published AON evaluated for their ability to induce or not exon-skipping in human cells and animal models. We compared the efficiency of these AON and their SKIP-e specificity predictions. We observed that the most active AON have a higher specificity vice versa.

NR-Analyzer: If exon-skipping could become a therapy of choice for some proteins, it is probably limited to specific transcripts...
only as the resulting truncated protein needs to retain enough activity to restore a function. Another genotype-based therapeutic approach has been proposed to induce premature termination codons (PTC) readthrough. This approach is theoretically applicable to all diseases for which a nonsense mutation is involved. Nevertheless, this assumption is incorrect as the readthrough will result in the random insertion of an amino acid resulting in a quasi-normal protein that contains an amino acid substitution.

RD-Connect partners therefore created the NR-Analyzer system allowing the annotation of all potential stop codons from any human coding transcript. This annotation identifies all potential Single Nucleotide Polymorphisms (SNP) that can result in the creation of a premature termination codon via a nonsense mutation. These SNPs may thus be selected to collect additional information including: the frequency of this mutation in the general population; the presence of missense mutations (pathogenic or not) at the same protein residue and their frequency in the general population; and a conservation score of the neo-proteins resulting from a potential nonsense read-through therapeutic strategy. Overall, the NR-Analyzer software allows the annotation of all potential stop codons from any human coding transcript. Thanks to its unique features, it allows to rapidly evaluate the potential of each nonsense mutation for a nonsense mutation read-through therapeutic strategy as exemplified by the analysis of CFTR gene mutations. In fact, the 3 most studied nonsense mutations E60X, G542X and R1162X correspond to very good candidates as illustrated in table 4.1.1 for the E60X mutation. More globally CF patients harbouring a nonsense mutation are good candidates for this therapeutic approach as 91.9% of them have a favourable context.

The Crawfish system: RD are caused by mutations in multiple genes (at least 7,000 RD are described and only 50% of the disease-causing genes are identified). In addition, most of known mutations are private (family specific) and only a small number is found in multiple families such as the c.Phe508del (∆F508) mutation of the CFTR gene involved in Cystic Fibrosis. In this context, the challenge for therapeutic strategies in the RD domain is to limit the mutation-specific therapies that will benefit only to few individuals and design new global approaches that could be used for multiple individuals. This implies to move from the gene-specific approach such as the exon-skipping strategy to broader approaches such as the PTC read-through strategy that targets nonsense mutations. Nevertheless, these strategies have various limitations:

- The exon-skipping approach will result in the production of truncated proteins that retain a certain level of activity but will never be fully active.
- The exon-skipping approach can only be applied to genes/proteins that tolerate internal deletions, which restrict its application to a handful of genes and mutations.
- The PTC read-through strategy will result in the production of a mix of proteins harbouring various missense mutations at the PTC site. Those proteins might either have a normal activity or an abnormal activity. This impact is difficult to anticipate even if the NR-Analyzer tool might help to select the most promising mutations.
- Both approaches will impact the two alleles of the target gene and therefore potentially damage the functional allele, especially in dominant diseases.
- In compound heterozygosity, where the 2 alleles must be rescued, those strategies might not be adapted.

A new therapeutic approach that may target a specific mutation or group of mutations and restore a fully functional protein without damaging the normal allele was therefore needed. A such approach has been proposed by Puttaraju et al. in 1999 and is called the Spliceosome-mediated RNA trans-splicing (SMART). This approach aims to repair RNA to restore a normal mRNA molecule. To do so, pre-trans-splicing molecules (PTMs) capable of base pairing to and trans-splicing with a conventional pre-mRNA target are used. They can induce 5’, 3’ or 5’+3’ trans-splicing. They are composed of a Binding Domain (BD), of a linker region to an acceptor or donor splice site and to the first exon(s) or last exon(s) to be replaced. The Crawfish system contains data from the Human Reference Genome HG19: 18,392 genes; 29,538 transcripts; 215,921 exons and 143,475,377 Binding domains of 153 bp.

The optimized pre-trans-splicing molecules (PTMs) were designed as a combination of the Binding Domain linked to an intronic region containing an optimal branch point (consensus-value of 100 according to Human Splicing finder) and an optimal donor (5’ trans splicing) or acceptor splice site (3’ trans splicing) according to the Human Splicing Finder matrices (http://www.umd.be/HSF3/). The optimal donor and acceptor splice sites were derived from the 2 first bases of the constitutive exon (acceptor site) or the last 3 bases of the exon donor site) with respectively the optimal last 12 bases or the first 6 bases of the artificial intron.

Impact of therapeutic modifications of gene expression in rare disease
Metabolic phenotyping/profiling is essential for rare disease translational and clinical research. Targeted and exploratory metabolic phenotyping allows for the analysis of patient and population-based samples for the purpose of novel biomarker, diagnostic and prognostic marker discovery.

The objective of this task was to implement and deploy an open source tool to allow the evaluation of the impact of therapeutic modifications of gene expression in rare diseases and streamline Metabolomics standardized analysis.

MASTR-MS is a web-based collaborative laboratory information management system (LIMS) for metabolomics. MASTR-MS is a comprehensive LIMS solution specifically designed for metabolomics. It captures the entire lifecycle of a sample starting from project and experiment design to sample analysis, data capture and storage. It acts as an electronic notebook, facilitating project management within a single laboratory or a multi-node collaborative environment. This software has been developed in close consultation with members of the metabolomics research community. It is used throughout Australia and international groups have been utilizing it. It is freely available under the GNU GPL v3 license and can be accessed from, https://muccg.github.io/mastr-ms/ for the RD Connect community.

This system has been fully described in D4.21. In short, it is a downloadable and installable LIMS solution that can be deployed either within a single laboratory or used to link workflows across a multisite network. MASTR-MS is a web-based LIMS solution for metabolomics laboratories. It comprises 5 major modules: (1) the Node Management System; (2) the User Management System; (3) the Quote Management System; (4) the Project Management System and (5) the Data Management System. The different modules allow users to:

- Track all metabolomics samples and associated meta-analytical and processed data sets. This starts from the capture of client/collaborator communication, the establishment of new projects, experimental design and sample definitions and the automatic capture of raw data generated by the instruments.
- Develop an electronic notebook, where users record all relevant information about projects and experiments in MASTR-MS, thus allowing multiple users to work on the same project.
- Methodically manage the vast amount of data generated by the analytical instruments, by associating it with the project, experiment and sample details.
- Facilitate collaboration between geographically distributed laboratories through the sharing of projects and experiment data.

Thanks to the high quality of the MASTR-MS system, it has been successfully deployed and utilised in Australia as part of the Metabolomics Australia. In parallel, there is strong interest to deploy in other laboratories around the world, e.g. Albert Einstein College of Medicine, Bronx, New York 10461.

Finally, as other systems from P4, it is available to the community at https://muccg.github.io/mastr-ms/.

3D facial analysis system

The face is a biological billboard that is uniquely representative of health and disease. It has been particularly investigated for clinical translation for diagnostics and treatment monitoring of rare, and often genetic, disease and other syndromes. Intrinsic diagnostic information resides in facial data, particularly when acquired with deep and 3-dimensional precision.

The Australian RD-Connect partner has developed methods to create facial diagnostic signatures of rare diseases that have been refined to detect subtle facial variations in a personalised and holistic manner, to explore disease biology in combination with text-based information, and to objectively monitor therapeutic response with deep precision.

The CliniFace system (http://www.crcsi.com.au/research/4-4-health/current-projects/4-412-cliniface/) aims to deliver a suite of software tools packaged within a single overarching application named 3D-FAST (3D Facial Analysis Streamlining for clinical Translation). It is interoperable with multiple 3D camera types, and various other imaging and video file types; that is modular and with expanding breadth and depth of new and legacy analytic modules; and has multiple file output structures for integration with multiple databases, registries and platforms.

The vision for 3D-FAST is to quickly and accurately analyse a 3D scan of a patient’s face (as captured by existing photogrammetric hardware) to provide a summary to a clinician. The result is a series of facial characteristics that are likely to suggest some underlying genetic condition. Thirty per cent of rare diseases patients wait up to 30 years for a diagnosis. Thirty per cent see six or more doctors before receiving a diagnosis and nearly fifty per cent receive an initial diagnosis that is incorrect. 3D-FAST is expected to significantly improve upon existing methods of automated facial analysis for assisting in
syndromic diagnosis, especially in the realm of rare disease diagnosis where there is limited clinical data. This stage of the project aims to answer two key research questions: Is it possible to accurately classify syndromes and Human Phenotype Ontology (standard for describing human variation) terms from 3D scans of a patient’s face? Are there inferential associations between a patient’s genetic information and their facial (dys)morphology (abnormal features)? In order to answer these questions researchers will extend the 3D-FAST capabilities to include several critical features such as facial co-registration, facial averaging, analysis of facial differences and symmetry, and detection and classification of salient facial morphological characteristics.

This innovative research has led to real-time data mining for comparisons with a repository of facial imagery for powerful diagnostic and treatment monitoring. In time, it will significantly improve clinical efficiency and patient outcomes. The next development phase of CliniFace includes growing a database of facial imagery that clinicians can utilise and compare captured faces and facial landmarks against normalised faces in determining disease types. Future collaborations include the Fiona Wood Burns Unit and research into Down’s Syndrome and Foetal Alcohol Syndrome.

The latest version of the open source tool has been provided to the RD-Connect platform. There are a range of enhancements, including an export function that generates a package of all required files for visualisation along with meta data and the first iteration of embedded HPO term generation.

The tool is also currently being integrated with multi-omics platforms in Australia and Japan, and is also a named tool in a pan-Canadian rare diseases initiative. It is also being used by an increasing range of clinical services in metropolitan and rural and remote regions.

Most recently, through a competitive grant process, the WA Department of Health, in partnership with Curtin University, Perth Children’s Hospital and the premiere clinical trials unit in the Southern Hemisphere, Linear Clinical Research, has funded a body of work. Specifically, to accelerate facial biomarker functions of 3D facial analysis for existing and trial rare diseases treatments; Facing Clinical Trials.

Harmonisation of RD-Connect tools with international activities on electronic health records

While full integration of data from Electronic Health Records (EHRs) is beyond the scope of RD-Connect owing to the tremendous challenges involved, RD-Connect continues to be active in initiatives where this topic is of relevance. In particular, the substantial activity relating to European Reference Networks has resulted in a lot of activity related to the sharing of health data internationally for healthcare reasons while also enabling its reuse for research. RD-Connect WP4 partners from LUMC actively support ERNs in making their data FAIR and aim to ensure these concepts make their way into EHR developments. Work with EHRs is ongoing, for example the Personal Health Train initiative in the Netherlands (http://www.dtls.nl/fair-data/personal-health-train/) which aims to bring research to the data rather than the classical solution of bringing data to the research, enabling access while ensuring maximum privacy protection and maximum engagement of individual patients or citizens.

AMU organized an Electronic Health Record (EHR) workshop that took place in Marseille on July 3rd, 2014, to investigate the possibility to extract information from such systems. It gathered 20 participants from various backgrounds including key partners from the Electronic Health Records for Clinical Research (EHR4CR) network, which reported on the project achievements including the creation of the EHR4CR Institute. The workshop participants recognised that the key technical challenges were connectivity and interoperability, compliance with ethical, legal and privacy requirements from different countries and quality assurance of the data. They also concluded that even though EHR might have strong potential for information extraction related to Rare Diseases, many issues should be solved to make it a reality.

Ethical, legal and social issues

Throughout the project, ethical, legal and social issues in genomics and biobanking have been actively explored by RD-Connect partners under the leadership of Mats Hansson (UU). This proactive approach to ELSI activities and full involvement of patients has ensured due consideration of these important aspects. As part of the work done on the regulatory aspects of research, UU and EURORDIS have also represented RD-Connect in the work led by BBMRI-ERIC on the Code of Conduct for Health Research, aimed at providing useful guidance on the application of the EU GDPR. RD-Connect participates in the writing group for the code of conduct and is leading the taskforce on Informed Consent. In January 2018, UU, in collaboration with EURORDIS and BBMRI ERIC organized a workshop on risk based ethical review, hosted by EURORDIS in Paris. UU also led the work to adapt the processes and documentation of the RD-Connect GPAP (including its Code of Conduct and the Adherence...
Agreement) to the EU GDPR.

Identify regulatory hurdles

WP6 work started from the ongoing identification and analysis of regulatory hurdles, as task conducted throughout the project and especially developed in the first year of the RD-Connect. Led by UU in collaboration with most partners of RD-Connect as well as with scientists and ethicists in other European networks, this work started in 2013 with the analysis of the GDPR proposal from the Committee of Civil Liberties, Justice and Home Affairs of the European Parliament. This analysis led to the conclusion that, if the newly proposed Personal Data Protection Law as it was would come into effect, this would have had severe repercussions on the possibilities to do biobank- and registry based medical research. Based on our analysis we triggered public opinion and started a conversation in Nature Reviews Genetics, mirroring the worries by many other scientists that lead to profound changes in the proposal and in the law. (Mascalzoni D, Knopper BM, Aymé S, Macirotti M, Dawkins H, Woods S, Hansson MG, Rare diseases and now rare data?, Nat Rev Gen 2013, doi:10.1038/nrg3494).

The duty to maximize the use of precious and rare resources in rare diseases makes data-sharing one of the most compelling problems facing the progression of research. Regulatory burdens often hamper a free flow of data. A stakeholder conference was held in October 2013 in Brussels, hosted by UU, in collaboration with EURORDIS and UNEW-PEALS. Patients representatives were invited by EURORDIS to contribute to the debate to identify regulatory hurdles for international data-sharing in rare diseases research specifically the ones related to informed consent, trying to look for shared solutions. A comprehensive and shared regulation does not exist yet.

In conjunction with the European Academy of Bolzano and the EU COST Action CHIP Me, UU organized an international workshop on Data Sharing in International Databases with Focus on EU directive and Across Boarders Sharing (Genetic data in public research databases: Which governance mechanisms should apply? April 27-28, 2016, EURAC Bolzano Italy). The workshop explored ethical and legal challenges that arise when researchers are required to deposit genetic and genomic research data in public research databases as well as investigate governance mechanisms that may support ethically and legally compliant data deposit. Experts from Europe (including representative from the EU Commission), USA, Canada and Australia joined the discussion and a position paper is under review in Annals of Internal Medicine.

Public private partnership in research constitute a great challenge for research on the ELSI perspective, including issues with regard to conflicts of interest and distrust from the public and from the patients. Some barriers to effective PPPs include a generalized lack of common regulatory frameworks which burden the potential positive impacts on citizens wellbeing. On the 7-8 November 2016, Uppsala University organized and hosted the workshop “Ethically and legally sustainable partnership between industry and public funded research initiatives: PPP and Rare diseases as a case study” co-branded with RD-Connect and CHIPME with 50 participants from across Europe. The workshop was attended by members of the RD-Connect team, PEC representatives and RD-Connect SAB. The workshop focused on identifying conditions for a successful Public private partnership. Both the programme and the report have been widely disseminated and circulated among partners. There Stakeholders from different areas presented cases of public private partnerships (PPP) in research, looking at ethically sustainable partnerships with the patients in mind. The most commonly used models are not the most effective or economically viable. Current methodology focuses on addressing conflict of interest but not maximising output or addressing the needs of all stakeholders. A ten point benchmarking assessment has been published in the dedicated deliverable available online, as a guideline for future PPP, this contains things that can be assessed and measured to determine the partnership. This is the first-time ethical principles have been transferred into a legal framework that we are aware of.

As part of the work done on the regulatory aspects of research, UU is currently representing RD Connect in the work lead by BBMRI ERIC on the Code of Conduct for Health Research, aimed at providing useful guidance on the application of the EU GDPR. RD –Connect is full part of the Writing group that meet several times last year and is leading the taskforce on Informed Consent. The purpose of the Code of Conduct for Health Research initiative is to contribute to the proper application of the regulation, taking into account the specific features of processing personal data in the area of health; To clarify and specify certain rules of the GDPR for controllers who process personal data for purposes of scientific research in the area of health; To help demonstrate compliance by controllers and processors with the regulation; and to help foster transparency and trust in the use of personal data in the area of health research.

Develop guidelines for informed consent in prospective studies

As part of the ELSI work in the RD-Connect, WP6 ran a substantial analysis of the informed consent requirements for the
platform along with the development of guidelines for practical implementation. A first deliverable looked at consents used by partners. A stakeholder workshop on Informed consent in RD-Connect was then organised and held in ISS in Rome, on April 23-24, 2014. The workshop aimed at discussing and collecting the opinion of relevant stakeholders (representatives of the scientific community coming from B-projects and RD-Connect, patient representatives, scientists and bioethicists involved in the preparation of guidelines on informed consent in other international consortia) to agree on common values and procedures prior to drafting Guidelines on informed consent in RD-Connect. Specific attention was paid to the old collections obtained with / without informed consent and of course the conditions for new collections. Following the workshop a first draft of the guidelines for informed consent in RD Connect were circulated amongst all partners collaborating with WP6 to receive more inputs by all stakeholders and the final version of the guidelines is the result of such collaborative effort published by Sabina Gainotti in the EJHG and endorsed by IRDIRC. In addition, RD-Connect contracted the services of a consultancy firm to externally evaluate compliance with the EU GDPR and to generate standard informed consent templates as this is a recurrent request from data submitters. These templates are being generated in a way that links in with the informed consent templates produced for patients participating in the activities of the European Reference Networks, thus ensuring a harmonised experience for the patient and the doctor.

Develop an ethical framework for the platform

The ethical framework is meant to provide a basis to ensure uniformity of access across projects and countries, and may be regarded as a consistent basic agreement for addressing data and sample sharing in RD-Connect. As part of the ethical framework a discussion on ethical concerns for sharing was planned. The first milestone of the model has been the Charter for Sharing, to guide the sharing activities of the consortia, endorsed by IRDIRC, BBMRI.se and BBMRI.it. The Charter is the result of a careful negotiation of different stakeholders’ interest and is built on earlier consensus documents and on intense discussion among the partners. (Mascalzoni et al., EJHG 2015).

The activity contributed to the ethical framework by providing the state of the art in the discussion on return of information and results (RoR). In the report there is special consideration for issues related to RD patients. The report identifies and maps out the different scenarios in which the return of results may occur in Rare Disease Research. By closely looking at the definitions of what we understand under the categories of results, including incidental findings, we describe and discuss herein the different types of results that could possibly be conveyed to patients taking part in rare disease research projects. We also address the principles outlined in the literature that may help us understand how to deal with all these different types of results and if the return of results to rare disease patients may constitute a special case scenario (Viberg et al. EJHG).

A Report on involvement of children in longitudinal research looked extensively into ethical and legal approaches for the involvement of children in longitudinal research. The report addresses key concerns for the involvement of children in longitudinal research through a literature analysis providing sources and an overview of legal and ethical obligations with regard to assent, consent and re-contact when reaching adulthood. Moreover, the report outlines the different stakeholder voices collected during the Workshop on Children’s Involvement in Research held in September 2015 23th and 24th at Uppsala University, in order to discuss relevant issues regarding children involvement in longitudinal studies. The focus of the workshop was on assent and consent. The involvement of data submitting partners added strength to the workshop, along with the involvement of experts in consenting children. Children and parents with rare diseases, currently involved in research participated in the workshop via videolink. Selected literature has been circulated in order to prepare the workshop discussion background.

The report will be further developed into a joint paper, that will include the relevant stakeholders in order to help dissemination of the results. Further work would be required on this specific topic. Open questions still remain also in the first stage report on pre-manifesting carrier status that provides an overview of the current literature on normative practices about the communication of carrier status in families involved in research. The relation of this topic with unexpected and secondary findings becomes very clear when we analyse current practices. There is a divide between clinical and research attitudes towards carrier status communication. In fact, where genetic counselling in the clinical setting follows clear directive for the involvement of family members it is still unclear what the current practices are in the research setting. Guidelines, based on current literature and regulations are provided, but the authors feel that empirical research to collect the inputs from the stakeholders is needed, but further research is needed in the field to account for patients preferences.

Also a central piece of the ethical framework has been developed in the form of a Code of Practice (later renamed a Code of
Conduct) regulating access and use of the RD-Connect data platform. The Code has been circulated to and commented on, by scientists, lawyers and ethicists belonging to RD-Connect, NeurOmics and EURenOmics. It has also been reviewed by patient representatives and EURORDIS as well as being presented and discussed at a meeting with the Section for Ethics and Integrity at the European Commission. An adherence agreement has also been developed to be signed by the users of the platform. The Code of Practice and the Adherence Agreement were approved at the Executive Management Committee of RD-Connect 11 June 2015.

Risk-based model for ethical review analysis

The paper “Patients would benefit from simplified ethical review and consent procedure (Hansson MGet al, Lancet Oncology 2013)” has been the basis for considering risk based ethical review. This approach is embedded in the whole policy developed for the consortia, the code of practice, the indications on how to treat data and samples.

It is also reflected in all our publications with special emphasis on how to share data and samples from existing and perspectives collections (Mascalzoni 2015, EJHG) and how to ask consent by assessing the presence of risky “items” in the research project (Gainotti 2016, EJHG). Those last two papers have been awarded the “IRDIRC recommended” and indeed have been developed together with the stakeholders in order to provide also patient’s insight into the recommendations. Both papers provide “checklists” that can be useful for ethical review for the use of data and samples in research and the Sharing charter has been officially endorsed by BBMRI. SE.

In January 2018, UU, in collaboration with EURORDIS and BBMRI ERIC organized a workshop on risk based ethical review, hosted by EURORDIS In Paris. The aim of the workshop reflects the work developed during the whole RD-Connect project and is built on the work done on the regulatory aspects, including consent and on the work done with patients. The premise is the importance of a governance framework and of ethical review for biobank-based research projects, that may be built on different premises than clinical work in order to avoid unnecessary hurdles. The workshop was attended by several ethicists and bioethicists across Europe and EURORDIS presented the work carried out within RD-Connect and provided an overview with regards to expectations of rare disease patients relating to the sharing of their data, as well concerns about the GDPR. Patients recognise the importance of data sharing but view consent as necessary so that patient preferences can be respected. Progress and outcome of research should also be communicated with patients. Furthermore, protection of privacy and confidentiality is critical. Trust and transparency are key, as well as critical overview.

The results show that the growing complexity in the use of existing data highlights some gaps in the Ethical Review model currently in use across Europe. Reports form scientists highlight that the typology of reviews and constrains applied throughout Europe are highly uneven.

Research using human data and samples should occur under ethical scrutiny. The general rule is applied unevenly across borders and even across institutions. While some Review Ethical Committees (REC) apply strict constrains for primary and secondary uses of data and samples by binding possible projects to strict and specific consent, others do not apply the same rules to primary and secondary studies. In certain instances, if data and samples are used for secondary studies, no ethical review is performed at all. There is a compelling need to push the conversation towards agreed standards and minimal requirements. A report developed with BBMRI ERIC will be ready for publication in the following months.

Engage with patient groups and patient organisations

As carried out in close collaboration with the patient-led activities, patient involvement and engagement in RD-Connect was substantial and reflected in all the activities presented. The Rare Disease Patient and Ethics Council (RD-PEC) was formed to ensure close collaboration on common problems related to ethical, legal and social issues (ELSI) which arose out of the work RD-Connect, Neuromics and EURenOmics. It is comprised of the members of the Ethics Advisory Boards from the three projects, representatives from the Patient Advisory Council and representatives of the RD-Connect ELSI workpackage and patient engagement workpackage. The RD-PEC is a high level advisory body examining ethical, legal, social and participatory issues linked to research taking place in the context of the 3-projects.

UNEW-PEALS in consultation with the RD-PEC and PAC continued and completed the work on the inclusion of patient organisations in platform governance. Multiple models of governance were debated and refined over a number of months before being discussed in a telephone conference with PAC members where a model was settled upon. This was documented by UNEW-PEALS and presented to and accepted by the EMC. It was noted that a priority for future governance must the continued inclusion of patient representatives.
At the RD Connect meeting in Athens, 16-18 April 2018, Pauline McCormack (UNEW-PEALS) hosted a workshop Patient Involvement in Future Project Impacts with patient representatives and EURORDIS. This discussed and prioritised topics of interest to patients at this stage of the project and after completion of EU funded phase. This was co-led with patient representative Veronica Popa and participants reported back to the wider meeting on topics discussed.

Patient involvement

Work relating to patient involvement has been led by Virginie Bros-Facer on behalf of EURORDIS and has strongly confirmed the added value of patient engagement in a research infrastructure project. The benefits of having rare disease patient representatives involved as a group to discuss and exchange on issues of particular interest to patients, as well as having dedicated patient representatives directly integrated in the different work packages, have enabled the collective rare disease patient perspective together with the specific and individual input to be integrated as and when needed. These patient-led activities are being continued in future through work within the Solve-RD project and the EJP-RD.

Patient-led activities were carried out in close collaboration with the ELSI work from the very start of the project and joint activities included the active participation of the Patient Advisory Council within the Patient and Ethics Council and focus groups dedicated to gathering views and perspectives of patient representatives on large-scale data sharing of genomic data enabling the identification of priorities which served as the basis of empirical research on this topic. A range of issues were identified: what can we learn from historical precedents; the responsibilities of researchers; who should have access to collections and how this should be managed; and how researchers can improve communication with participants (McCormack et al; 2016). Capacity building workshops were also jointly arranged to prepare patient representatives for full participation in scientific meetings and this ensured the participation of several patient representatives from a range of rare disease communities for true patient engagement. Topics discussed included: i) the importance of -omics research and bioinformatics, ii) data integration system of RD Connect, iii) a Global Unique Identifier for Rare Diseases, iv) informed Consent in the context of the European Data Protection Legislation, v) policy on incidental findings and vi) data access and commercialisation of future results. The work related to the Global Unique Identifier for Rare Diseases was later taken up by a dedicated Task Force initiated by RD-Connect and co-organised by the Global Alliance for Genomic Health and IRDiRC in which EURORDIS participated.

In addition, EURORDIS has facilitated discussions and organised workshops with the PAC and the wider rare disease patient community on ELSI topics relevant to the project activities. The conclusions of these discussions constituted an important part of several published recommendations including developing standards and guidelines for informed consent (Gainotti et al; 2016), the International Charter of principles for sharing bio-specimens and data (Mascalzoni D et al; 2014). Both of these guidelines have received the label IRDiRC Recognized Resources. In January 2018, EURORDIS hosted at its offices in Paris and co-organised together with RD-Connect partners, BBMRI-ERIC and Uppsala University, a workshop to reflect on the importance and needs of ethical reviews for biobank-based research projects. The workshop was attended by several ethicists and bioethicists across Europe and EURORDIS presented the work carried out within RD-Connect and provided an overview with regards to expectations of rare disease patients relating to the sharing of their data, as well concerns about the GDPR (collected via a webinar with EURORDIS members and follow up discussions with the PAC). Patients recognise the importance of data sharing but view consent as necessary so that patient preferences can be respected. Progress and outcomes of research should also be communicated with patients. Furthermore, protection of privacy and confidentiality is critical, and trust and transparency are key.

Integrating several PAC members within the activities of the ‘technical’ work packages successfully enabled the direct contribution of patient representatives and consequent appreciation of the added value that patient engagement brings to RD research activities. The PAC provided significant contribution to the quality self-assessment check list for registries led by WP2 which was presented (Yilka Kodra) as a poster during the annual RD-Connect meeting in Berlin in May 2017 and published EURORDIS participated as a speaker in the 3rd, 4th and 5th International Summer School on “Rare Disease and Orphan Drug Registries” organised by ISS, Rome where engagement of patients in registries and the different ERNs through EURORDIS support activities to the European Patient Advocacy Groups (ePAGs).

One member of the Patient Advisory Council of RD-Connect supported by EURORDIS (Marieke van Meel) is a member of the Biobank Assessment Panel led by WP3 (FTELE) and has contributed to the evaluation of candidate biobanks to join the RD-Connect sample catalogue. In addition, EURORDIS continued to support and liaise activities between the PAC members and
WP3.

EURORDIS as a member of the BBMRI Stakeholder Forum is a member of the large consultation group for the development of a code of conduct to help scientist comply with the GDPR. During the BBMRI-ERIC Forum Meeting on the development of the Code of Conduct and the BBMRI Stakeholder forum meetings organised in 2017 and 2018, EURORDIS communicated RD patients’ main concerns and outstanding questions regarding the GDPR highlighting the need to clarify issues surround secondary use of data, data of deceased persons as well as ethical issues related to genetic data.

A subgroup “the editorial board” was created for the newly developed specific section for patients and families on the RD-Connect website. This editorial board comprises highly motivated PAC members who are dedicated to further communicate and disseminate RD-Connect activities and outputs especially regarding bioinformatics, registries, biobanks and data sharing to improve RD research. This group of PAC members has written a series of 9 articles available on the RD-Connect website and disseminated through the RD-Connect monthly newsletters. A short series of video testimonials has also been prepared with PAC members and WP7 and are available on the RD-Connect website. Furthermore, a short series of infographics prepared by WP8 highlighting the core recommendations to improve qualities of RD registries will be made available on the RD-Connect website and disseminated through the various EURORDIS channels before the end of the funded period.

Potential Impact:

Impact on the rare disease research environment

The wide use of the RD-Connect infrastructure is a key indicator for its impact on rare disease research, diagnosis and healthcare. In six years, RD-Connect has become an important player in the rare disease field. The RD-Connect infrastructure has gained visibility and is already being used by many different projects and stakeholders. Important milestones in the use of the infrastructure were the successful collaborations with the major data submitters: NeurOmics, EURenOmics, EuroBioBank, 17 BBMRI-LPC projects, Solve-RD and the European Reference Networks. By October 2018, the GPAP had received 4000 genomic and phenotypic datasets, the Registry & Biobank Finder had recruited 360 rare disease patient registries and 22 biobanks, and the Sample Catalogue had received the data of over 250,000 rare disease biosamples. Through these collaborations, the project has jointly contributed to the discovery of over 100 novel disease genes. This number is expected to increase in the near future since RD-Connect serves as the primary analysis tool for the 17 BBMRI-LPC projects and is also being used by Solve-RD to diagnose unsolved cases from the European Reference Networks.

Each specific work stream had many impacts on the RD research environment.

User-friendly diagnostics and gene discovery

The GPAP has contributed to the RD research field by providing means for broader data sharing and analysis, allowing the identification of new causative variants and genes. This has a clear impact on patients which have received a diagnosis that had eluded previous analysis. Furthermore, it enhances overall RD knowledge, facilitating the diagnosis of future patients and enhancing the understanding of these diseases at the molecular level, thus opening new doors for the development of potential treatments. At the time of writing, 586 users representing 207 organizations are registered on the Genome-Phenome Analysis Platform.

As envisaged when the analysis system was initially set up, the GPAP is thus fulfilling the dual role of enabling data sharing while also lowering the barriers for rare disease researchers to analyse the data they submit themselves even without bioinformatics expertise. Reluctance on the part of researchers to share data due to concerns about losing ownership and being scooped in gene discovery publications was an early hurdle that RD-Connect needed to overcome. The recognition that a mechanism needed to be developed to allow the submitting clinicians and researchers to retain control over their data by empowering them to analyse it themselves was a major success. As has already been shown by the success of the European BBMRI-LPC project (900 exomes) and others such as Consequitur (UK-Turkey collaboration, 500 exomes) which submit data to the system, clinical academics enthusiastically use the system to analyse the patients they themselves have submitted, and this has brought tangible results including gene discovery publications and diagnoses for patients who were previously undiagnosed. The inclusion of even larger amounts of data from new projects such as Solve-RD will further increase this impact.

A catalogue to give visibility to patient registries

The Registry & Biobank finder is a unique catalogue-style resource that enables the owners of registries and biobanks to publish information about their resource. This increases the visibility of these important resources and provides a mechanism
for researchers, particularly those from European Reference Networks, to locate the data and samples they require. This will have an impact on the ability of researchers to do their research, lowering the barriers to locating relevant data.

Quality standards and data linkage plan for registries

A further important achievement was the development of recommendations for registry quality. The recommendations developed by the expert working group are to be used as a framework for improving the quality of RD registries, and this will have an impact on the standards of registries in Europe.

Even more crucially, the strong focus on Findable, Accessible, Interoperable and Reusable (FAIR) data in the registries community is having a substantial impact on the RD field as a whole. It is increasingly recognised that these mechanisms are essential to develop further in order to maximise the value of already generated data, and this is a major aspect of new projects such as the EJP-RD.

There remains a need for coordination between ongoing registry-related initiatives at national and international levels. At national level, we should recommend the development of centralised, public, national, “registry-as-a-service” platforms; centrally resourced platforms will guarantee the access to highly trained staff on the quality of the registry, foster the standardisation of the registries, allow economy of costs and time for setting up new registries, allow to interlink key data sources on different diseases, increase the capacity to develop cooperation at EU level. At a European level, the national platforms for RD registries are collaborating together for creating a centralised European Union-wide framework on patient registries with the benefits of data sharing, reducing duplication of efforts and costs, facilitating validation of results, enabling engagement with experts and the patient community and overcoming the “rare disease problem” in terms of cohort size, powering trials and finding confirmatory cases.

A working community of biobanks

Rare disease biobanks are crucial infrastructures for diagnosis and research. The availability of biospecimens and associated data of patients affected by RD plays a pivotal role in the identification of disease genes and molecular biomarkers, and in the development of novel treatments. RD-Connect developed a comprehensive platform of tools for making human RD biomaterials and their linked data accessible and available to the scientific community. The Sample Catalogue and Finder tools enabled RD biobanks to join forces to share resources and become an integral part of research infrastructure for RD. We have demonstrated real partnerships between researchers, biobanks, sequencing services, bioinformaticians, analyses and data sharing platform via the BBMRI-LPC research call, where a successful working ecosystem of infrastructures and experts were in place to promote RD research.

Patient engagement in biobanking

Biobanks are encouraged to establish collaborations for collections and dissemination of biological samples. Collaboration in research Calls may be an effective way to ensure biobanking services are available to researchers and clinicians in awarded projects. Similarly, formal collaboration with patient organisations allows precious samples to be collected more systematically from patients and family members, creating critical mass in sample collections for research. The successful collaboration model of TNGB with patient organisations for sample collection especially engaged patients as a key stakeholder in RD research. Another example of patient engagement in WP3 was the representation of an expert patient representative in the Panel for Biobank Assessment. The patient representative can provide advice, stimulate discussions on biobank operations, and facilitate understanding of mutual goals. Such examples have a high impact in stimulating dialogue between biobanks with the patient community as well as fostering their direct engagement in research.

Standards and culture for sharing data and samples from biobanks

A streamlined biobank operation process, spanning from the collection and storage of high quality RD samples/clinical data to their wide distribution, can maximise the impact of biobanks for RD research. RD-Connect designed and facilitated the appropriate usage of RD biomaterials with a recommended workflow that has implications on best practices in sample and data sharing. In addition, sharing of metadata on precious sample collections to a centralised catalogue was reinforced within RD-Connect and EuroBioBank, which contributes to the culture of sharing and FAIR (Findable, Accessible, Interoperable, Reusable) principles. These standards and guides for data/sample sharing are valuable in ensuring biobanks and researchers can work efficiently together as well as building trust during the exchanges. Similarly, the ethical, legal, social implications and challenges on data/sample sharing are applicable to all biobanks. The contribution to the drafting of GDPR Code of Conduct for Health Research has vast implications on the definition a widely recognised code on how to share information in biomedical
research. The results and experience of RD-Connect are not only valid for RD research or biobanks, but are valid for other biobanks and research areas.

Innovative bioinformatics solutions

The innovative bioinformatics tools developed as part of RD-Connect will contribute to the achievement of the new IRDiRC goals: 1) All patients coming to medical attention with a suspected rare disease will be diagnosed within one year if their disorder is known in the medical literature; all currently undiagnosable individuals will enter a globally coordinated diagnostic and research pipeline; 2) 1000 new therapies for rare diseases will be approved, the majority of which will focus on diseases without approved options.

The tools (UMD-Predictor, Human Splicing Finder, VarAFT, ALFA, YABI) to handle sequencing data (genes panels, WES or WGS) can aid in pinpointing disease-causing mutations that can then be experimentally confirmed. These systems can be linked to any existing bioinformatics pipeline and have been included in the RD-Connect Genome-Phenome Analysis platform for research purposes. Clearly, this had a strong socio-economic impact as it significantly contributes to the identification of new disease-causing genes and to reduction of the diagnostic odyssey. In addition, for the most complex situations where the disease-causing gene is unknown, RD-Connect partners released new guidelines and systems (LWAS, WGCNA, YABI, MASTR-MS) to be able to combine multi-omics data (genomics, proteomics, transcriptomics, metabolomics) into a single analysis. This approach also demonstrated its efficiency in various situations and will strongly contribute to the identification of new disease-causing genes. These new systems will strongly benefit to the general public as about 50% of RD genes remains unknown and are responsible for diagnostic wandering.

Because data are numerous but still scattered and difficult to manipulate, WP4 partners strongly supported the FAIR initiative to ensure proper data definition and interoperability, generation of semantic fingerprints for reference -omics and clinical phenotypes. These concepts were successfully translated into reality as illustrated by the SCALEUS (Semantic Web Services Integration for Biomedical Applications) system. While the impact of such initiative might not be directly perceived by the general public, it is a key element for the future to avoid silos and ensure efficient international collaborations to speed-up research.

Clinical diagnosis was aided through an innovative tool for 3D facial analysis. The CliniFace system aims to answer two key research questions: Is it possible to accurately classify syndromes and Human Phenotype Ontology terms from 3D scans of a patient’s face? Are there inferential associations between a patient’s genetic information and their facial (dys)morphology (abnormal features)?

Selection of appropriate treatment options was aided through two complementary approaches: The selection of the best therapeutic approach based on the genetic profile of the patients; and the creation of new therapeutic molecules. The ePGA system is a ‘one stop shop’ Web-based platform to ease the processing, assimilation and sharing of PGx knowledge, and facilitate the aggregation of different PGx stakeholders’ perspectives. The platform offers personalized diagnostics based on reliable genomic/genetic evidence and reduces healthcare costs by increasing drug efficacy and minimizing adverse drug reactions. The second aspect allowed the creation of innovative systems (SKIP-e, NR-Analyzer and Crawfish) that provide genome-wide information to assist researchers and drug companies to efficiently design new molecules to induce exon-skipping (SKIP-e), nonsense readthrough (NR-Analyzer) of trans-splicing (Crawfish), which are three of the most promising approaches for new RD therapies.

Together these “therapeutic systems” will benefit the general population as they will provide easy access to efficient drug selection based on genetic knowledge, on one hand, and, on the other hand potentially facilitate the design of new drugs for RD patients.

ELSI impact and impact of patient involvement

The ethical and legal documents produced in RD-Connect represent the joint effort of the stakeholders involved in the project and are an example of co-produced regulation. This experience demonstrated the importance of the involvement and engagement of patients into the development of policies to achieve standards that reflects common values.

The ELSI leaders engaged in dissemination activities that include more than 30 presentations in scientific conferences, more than 15 publications in scientific journals and organized several public events in which topics were discussed with stakeholders. This has had a substantial impact on the ELSI discussion in the RD community and beyond. RD-Connect also participates in the BBMRI-ERIC Code of Conduct writing group. This group is developing a code of conduct for the
implementation of the General Data Protection Regulation (GDPR), which will have a major impact on the way it is implemented for biomedical research. We are ensuring that the patient is at the centre of this development and the Code is directed to scientists working with data in Europe.

The patient-focused activities within RD-Connect have had an impact on the wider patient community, who have been kept informed through dedicated dissemination activities. By presenting at international, European and national workshops, EURORDIS has ensured that all stakeholders interested in the field of RD research were kept abreast of the latest development of RD-Connect activities and the crucial role that patients played in these activities. Enthusiasm and motivation to participate in the Patient Advisory Council has been ensured by recruiting several new members to join and provide a continuous flow of diverse patient perspectives and this has reinvigorated the group to participate and discuss ethical issues. Frequent and regular communication with the different scientific partners has enabled development of collaborations and increased patient input in activities within RD-Connect. Participation of PAC “delegates” within the “technical” work enabled a direct interaction with the aim to reflect a more active role of the PAC within the project and helps to maintain patient engagement, increase transparency as well as improve patient representatives’ input further within the project and disseminate the projects’ activities to the wider patient community. For example, the substantial contribution of the PAC to the upcoming publication “Recommendations for improving the quality of rare disease registries” has demonstrated the added-value of patient engagement in academic publications by bringing much-needed expertise and experience on the topic. Capacity building to improve the knowledge levels of PAC members and the wider RD patient community via the organisation of workshops and webinars on ELSI and scientific issues related to the use of novel genomic technologies, data sharing, the Genome-Phenome Analysis Platform, data protection and the new General Data Protection Regulation have not only increased the impact of PAC members into the project’s outputs but also engaged a high number of patient organisations into the work and ethos of RD-Connect.

Furthermore, linking the activities of the PAC of RD-Connect with that of the patient-centred work in Solve-RD, with the European Advocacy Groups (ePAGs) of the European Reference Networks (ERNs) and with the upcoming EJP-RD will ensure de-duplication of efforts, increase dissemination of the outputs of the projects, improve capacity building for patient representatives, and ensure the longevity of the patient-centred work through its continuation in other projects.

Publications and dissemination activities
To reach a diverse audience, RD-Connect used a range of different dissemination strategies, such as scientific publications, presentations at international conferences, training activities, posters, flyers, articles in the press and online media. In total, RD-Connect has run over 1100 dissemination activities, recorded regularly and reported in the in the deliverable reports. Many of these activities were also announced to the rare disease community through social media and the RD-Connect website.

At the end of the sixth year of the project, there are around 490 publications related to RD-Connect, out of which 210 acknowledge RD-Connect and are listed on the project website. The number of publications acknowledging RD-Connect has been increasing each year (Fig 1). A list of RD-Connect peer-reviewed publications is displayed on the website https://rd-connect.eu/scientific-publications/. New publications are highlighted in the newsletter and social media. The major publications on RD-Connect resources include:


To reach a broader audience, RD-Connect has also engaged with the press. Several magazines and journals with different readership and in different countries have published interviews and articles about RD-Connect, including the Rare Revolution Magazine (https://goo.gl/Xxf1k8) Pan European Networks, Horizon (http://bit.ly/2jRs8QL) and more. The value of RD-Connect
has been recognised by the European Commission’s Directorate-General for Research, which has promoted RD-Connect as a success story on their website (https://goo.gl/Ymi9uS).

To ensure dissemination of project outputs to the scientific community, RD-Connect was presented in the form of posters, talks, workshops, booths and flyers at numerous national and international events focusing on rare diseases, specific diseases, genomics, and biobanking, including major conferences such as the European Society for Human Genetics (ESHG), American Society for Human Genetics (ASHG), and European Conference on Rare Diseases & Orphan Products (ECRD) and IRDiRC Conference. In addition, the work of RD-Connect was highlighted thanks to several prizes awarded to RD-Connect partners, including the Black Pearl Scientific Award for Lucia Monaco (2017) and Volunteer Awards for Chris Sotirelis (2018), JRC Malta Young Scientist Award for Joanna Vella and the ECRD Poster Prize for Dorota Badowska (2018). These activities raised interest in RD-Connect within the rare disease research community, which resulted in submission of large numbers of datasets submitted to the Genome-Phenome Analysis Platform, Registry & Biobank Finder and Sample Catalogue.

Fig. 1. RD-Connect scientific publications and dissemination activities.

RD-Connect organised six annual meetings: Sitges (2013), Heidelberg (2014), Palma (2015), Barcelona (2016), Berlin (2017) and Athens (2018), attended by over 100 people each year. The meetings brought together project partners as well as invited attendees from partner research projects and infrastructures, such as NeurOmics, EURLenomics, EuroBioBank, E-Rare, BBMRI-ERIC, Elixir and the European Reference Networks. The annual meetings provided opportunities to organise training sessions and workshops for external stakeholders and enhance the communication between partners from different work packages. The meeting in Berlin in May 2017 was held back-to-back with the final meetings of NeurOmics and EURLenomics. On 3 May 2017, the three projects hosted a joint open Outreach Day, attended by around 300 participants, including project partners, researchers, patient representatives, policy makers and industry representatives. In three sessions, the participants discussed key cross-cutting topics in rare disease research: data sharing, diagnostics and therapy. The event also included training sessions on the RD-Connect Genome-Phenome Analysis Platform and the Sample Catalogue.

Presence in online media

In its first year of operation, RD-Connect launched the official project website (rd-connect.eu) which has been a major channel for disseminating the project outputs and gave the project a clear identity. It contains information about the project and its outputs as well news (rdconnect.eu/news/) events (rd-connect.eu/events/training/) and downloadable training and dissemination materials, such as webinars, flyers, presentations, manuals, scientific publications and other relevant information. Every year, the website has attracted increasing numbers of visitors. In total, it has received over 120,000 visits from most countries around the world, with highest interest in the UK, USA, France, Italy and Spain. In September 2017, we re-launched the website with a new design and more user-friendly interface. Following that, we saw in average that users displayed more pages per session than before, suggesting increased in the content (Fig. 2). To enhance the engagement with the patient communities, we added the section “For patients and families” (rd-connect.eu/for-patients-and-families/) containing a video introduction and several articles by PAC members, explaining different aspects of rare disease research and the RD-Connect work. The website also contains a several dissemination materials, such as flyers for different stakeholders and presentations.

Fig. 2. RD-Connect website traffic.

After the re-launch of the website with a new, more user-friendly design, the number of page views doubled.

In 2013, the website was supplemented with the RD-Connect monthly newsletter (rd-connect.eu/newsletters/) which contained updates about RD-Connect activities as well as other news from the rare disease field. In total, we have published 46 issues of the newsletter. The interest in the newsletter has been increasing over years, with the highest number of 1458 subscribers in April 2018 (Fig. 3), which dropped after introducing the General Data Protection Regulation in May 2018. Despite of the reduced number of subscribers, we believe the mailing list is targeting the right audience, as since the reduction, the percentage of newsletter opens doubled.

To stimulate communication with stakeholders, particularly with research projects, infrastructures, networks and patient organisations, RD-Connect has been active on social media: Twitter (@ConnectRD, since 2015), YouTube (RD-Connect, since
Over the course of the project, hundreds of users have followed these communication channels, with the numbers increasing every month, including individuals as well as organisations, e.g. several ERNs, companies and patient organisations.

The visibility of RD-Connect in the online media has been significantly increased by the launch of the RD-Connect YouTube channel (https://www.youtube.com/channel/UCwwcUPJZfyWGaW13Lvao7Ag). A major impact on the popularity of the channel had the RD-Connect explanatory video (https://youtu.be/i0C03vqGhDM) which was released in May 2017 in 7 language versions and participated in the European Commission’s FP7 showcase. Other videos in the channel include webinars, video tutorials, interviews with diverse stakeholders. We have recorded thousands of views, with the highest numbers from the UK, Italy, Spain, Germany and France.

Fig. 3. Visibility of RD-Connect in online media. The interest in the RD-Connect newsletter, Twitter, Facebook and YouTube channels has been increasing across the duration of the project. The increase in the YouTube views was caused by the release of the RD-Connect promotional video in May 2017. The sharp drop of the number of the subscribers after April 2018 was related to the General Data Protection Regulation, which entered into force in May 2018. The emails of the subscribers who had not confirmed their subscription were erased from the system.

Collaborations

RD-Connect originally worked closely with two EU-funded research projects NeurOmics and EURenOmics, which applied and developed innovative omics approaches to advance research on rare neuromuscular, neurodegenerative and kidney diseases. The end-users from these projects uploaded the first datasets to the RD-Connect systems and provided valuable feedback that helped to tailor RD-Connect to the needs of rare disease researchers. RD-Connect partners have also supported NeurOmics and EURenOmics in data linkage and gene discovery and provided numerous training sessions at their annual meetings. In 2017, the three projects hosted a large joint final conference in Berlin to highlight the research successes of all three projects, and the joint Outreach Day, which provided an opportunity for an open, multi-stakeholder discussion on the key aspects of rare disease research.

After close collaboration since the start of the project, in year 4, RD-Connect established a formal partnership with EuroBioBank, which agreed to be the de facto biobank in RD-Connect. EuroBioBank members, in addition to RD-Connect partners and patient representatives, are members of the RD-Connect Panel for Biobank Assessment.

Throughout the project, RD-Connect interacted closely with the IRDiRC, with RD-Connect coordinator Hanns Lochmüller chairing the IRDiRC Interdisciplinary Scientific Committee and a number of RD-Connect partners contributing to all IRDiRC committees and engaging with various task forces, including the joint IRDiRC-GA4GH task force on privacy-preserving record linkage, which aims to enable the linking of datasets on the same individual across different databases without revealing the individual’s identity. RD-Connect has contributed to the IRDiRC goals of 200 new therapies and the means to diagnose most rare diseases by the year 2020, the first of which has already been achieved ahead of time, in early 2017. Currently, the project is participating in the efforts to achieve the new, more ambitious IRDiRC goals 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, and to develop 1000 new therapies for rare diseases by 2020 (for details see http://www.irdirc.org/irdirc-goals-2017-2027-new-rare-disease-research-goals-for-the-next-decade/).

Linking up with related initiatives is crucial to the project’s impact, success and sustainability. In 2014, RD-Connect established key external collaborations with the biomedical research infrastructures on the European Strategy Forum on Research Infrastructures (ESFRI) roadmap, in particular ELIXIR, BBMRI-ERIC. As multinational research consortia with unlimited duration and a legal entity status in the EU, they are important partners to support future sustainability of RD-Connect and a valuable source of expertise in areas of relevance to RD-Connect. In 2016, BBMRI-ERIC became a full partner of RD-Connect, responsible for assessing sustainability options for RD-Connect’s biobanking resources. Since a large number of RD-Connect partners are also belong to ELIXIR, many activities relating to the rare disease use case, the interoperability and training were carried out jointly, which has been valuable for both sides. This included the joint work on linking registry and biobanks data and co-hosting the “Bring Your Own Data” workshops for researchers working in this area. In summer 2018, BBMRI-ERIC and ELIXIR became members of the RD-Connect Community and joined its Executive Committee.

RD-Connect partners were involved in several Horizon 2020 projects including EXCELERATE (ELIXIR), ADOPT-BBMRI-ERIC
Research and Innovation

(BBMRI-ERIC) and CORBEL (multiple RIs), which led to extensive interactions with RIs, especially in areas relating to data interoperability and mutually beneficial exchange of information and expertise. RD-Connect participates in the Global Alliance for Genomics and Health (GA4GH), particularly in the Clinical Genomics Working Group and the MatchMaker Exchange project, an initiative to encourage secure sharing of information on genetic variants without compromising patients' privacy. Integration of MatchMaker Exchange allows users learn whether other databases contain their variant of interest or genes with specific variants and contact the data owners to request more information.

To align its outputs with public health infrastructures, RD-Connect has actively engaged with participants in the EU public health sphere, such as the Commission Expert Group on Rare Diseases, the EXPAND project for health data interoperability, RD-Action, the key EU project involved in the development of the European Reference Networks (ERNs).

The inauguration of European Reference Networks (ERNs) in March 2017, international networks of centres of expertise for specific rare diseases, created an opportunity for RD-Connect to increase its engagement in the public health sphere. While ERNs have primarily a healthcare focus, they must also establish research goals. To support the ERNs in the research and diagnostic aspect of their work, RD-Connect provides its expertise regarding sharing research-related clinical, biosample and omics data joined the Task Force on Interoperable data-sharing within the framework of the operations of ERNs and will. In 2018, as part of the Solve-RD project, RD-Connect has started a pilot on re-analysis of existing exomes in four ERNs: GENTURIS, EURO-NMD, ITHAKA, and RND, Six other ERNs -ERK-NET, EYE, GENTURIS, ITHAKA, RITA and ENDO – are interested in participating in re-analysis after the end of the pilot. RD-Connect liaises with all 24 ERNs to host their research sequencing data for analysis and we expect that in longer term all ERNs will become its users.

Global impact

Rare diseases affect around 400 million people worldwide and can be a particular burden for those living in less favoured areas and low-resourced environments, where the means to provide diagnosis and treatment are often scarce. With the decreasing sequencing costs, the technique becomes affordable in more and more countries, and therefore the demand for training in data analysis and interpretation is increasing around the world. As RD-Connect is available to researchers and clinicians around the world free of charge, it may have a particularly big impact in low-resourced regions, where it is already used for clinical diagnosis outside the pure research setting. To target the key regions, RD-Connect established various collaborations and participated in regional and national conferences in numerous countries, including Central and Eastern Europe, Middle East, East Asia and Latin America.

Since 2014, the active engagement of RD-Connect partners in the activities of the global research network Genomic Medicine Alliance and the Golden Helix Foundation led to numerous joint research and educational activities stimulating genomic medicine research, yielding concrete improvements, research projects and partnerships in several countries, such as the collaboration with the SERBORDISInn project (http://serbordisinn.rs/) in Serbia. We particularly need to highlight the rare disease genomics-focused Golden Helix Symposia in Egypt, Malaysia, Serbia and United Arab Emirates, and Golden Helix Summer Schools in Greece.

RD-Connect collaborates with 3Gb-TEST, an EU project aiming to raise awareness of innovations in molecular testing among healthcare professionals to help implement diagnostic genome sequencing in Europe. Both projects have organized together courses on Next Generation Sequencing (NGS) in diagnostic setting in in Prague, Czech Republic (2015); Lisbon, Portugal (2016) and Ljubljana, Slovenia (2017), attended by many researchers and clinicians from the EU priority zones as well as from outside the EU. RD-Connect also collaborates with RD-Action, a project working on data and policies for rare diseases. In 2017, they organised the national Europlan Workshop in Prague, well attended by participants from Central and Eastern Europe.

RD-Connect is also liaising with the Initiative on Rare and Undiagnosed Diseases in Japan, to support supports Japanese researchers in learning how to FAIRify their patient registries data and link them to the data in international data infrastructures. Researchers from the Japan Agency for Medical Research and Development and other institutions several times have visited the RD-Connect coordination office and attended the RD-Connect annual meetings to discuss potential collaboration and sharing of expertise.

To facilitate our engagement with rare disease research communities around the world, we released the RD-Connect promotional video in 7 language versions: English, French, Spanish, Italian, German, Russian and Arabic, and with subtitles available in 48 languages.

The above-mentioned dissemination activities helped to raise the interest in the RD-Connect infrastructure in numerous
countries around the world. As a result, the RD-Connect GPAP has currently registered users from many non-Western European countries: Bulgaria, Canada, Croatia, Czech Republic, Egypt, Greece, Hungary, India, Iran, Israel, Latvia, Malta, Romania, Serbia, Slovenia, South Africa, Turkey, Ukraine and USA.

Sustainability
The resources generated by RD-Connect are not time-limited and need to be sustained beyond the initial FP7 funding. Sustainability objectives included consolidating RD-Connect’s position in the community as the leading resource for access to biosamples and patient registries and for analysis of genome-phenome datasets for diagnosis and gene discovery, expanding collaborations with new data submitters and with European and global initiatives for rare disease research, and ensuring the sustainability of the RD-Connect resources through new sources of funding. We have successfully ensured that the resources developed by RD-Connect will continue to grow and develop thanks to a range of new funding and sustainability mechanisms, in particular the Horizon 2020 project Solve-RD, the upcoming European Joint Programme for Rare Diseases, collaboration with the ESFRI research infrastructures ELIXIR and BBMRI-ERIC, and the launch of the RD-Connect Community.

Beyond the resources themselves, one of the most significant achievements of the RD-Connect project as a whole was the creation of a unique, multidisciplinary community bringing together experts from different fields to work together on enhancing rare disease research and data sharing. To continue this work, the RD-Connect Community was officially launched in July 2018 as an independent, non-governmental, not-for-profit, international association of individuals and organizations sharing the vision of building an open community that works to improve rare disease research. Its mission is to promote, facilitate and accelerate rare disease research by maximizing the availability and (re)use of rare disease data and biosamples through provision of infrastructure, tools and services to share, analyse and link datasets and biosamples in a secure and regulated way. By promoting data sharing and analysis tools and the data sharing ethos among rare disease researchers and clinicians and by raising awareness among patient communities and policy makers, the community aims to maximise the impact of the tools and services developed by RD-Connect members. Interest in the Community is high – in the first three months of its operation, the Community received membership requests from 127 individuals, academic research groups, organisations and European Reference Networks from 29 countries in all continents.

Achieving sustainability through inclusion of the RD-Connect tools in new projects has satisfied the European Commission’s goal of building on existing infrastructure and capitalising on previous investment. This was in fact clearly mandated for one of the major initiatives to be approved recently, the European Joint Programme for Rare Diseases. This major initiative brings together more than 130 entities from 35 countries (including 27 EU Member States, seven Associated Countries and Canada) and has a total budget of over €100 million. It will formally launch in January 2019 and all the major RD-Connect resources play important roles. This includes not only outputs such as the Genome-Phenome Analysis Platform (GPAP), Registry & Biobank Finder, Sample Catalogue and bioinformatics tools, but also the data stewardship expertise and knowledge-transfer activities that grew up within the RD-Connect environment, including the application of the FAIR (Findable, Accessible, Interoperable and Reusable) principles to rare disease data and the many training activities such as the summer school for registries and biobanks and the training for patient representatives facilitated by RD-Connect.

In addition to the EJP-RD, the launch in January 2018 of Solve-RD, a major new Horizon 2020 project in which the RD-Connect platform plays a central role, has been an important new development. This project will see the submission of 19,000 new datasets to the RD-Connect Genome-Phenome Analysis Platform, which will be analysed by researchers from the European Reference Networks for rare neurological disorders, neuromuscular disorders, intellectual disability and genetic tumour risk syndromes. The continued work with the European Reference Networks, now including data submission through Solve-RD, is another highlight: these important international initiatives are now a major feature of the RD landscape in Europe and many of them are looking to RD-Connect to provide not only the infrastructure for omics data sharing and analysis but also the expertise in data stewardship including ontologies and FAIR data principles. In addition, the ESFRI research infrastructures ELIXIR and BBMRI-ERIC have become even more closely associated with RD-Connect, with BBMRI-ERIC committed to working with RD-Connect on the sample catalogue and registry and biobank finder, and ELIXIR incorporating RD-Connect data stewardship and data analysis activities under the banner of its Rare Disease community within the ELIXIR Human Data community.

RD-Connect operates within the context of the International Rare Diseases Research Consortium (IRDiRC) as one of the EU’s flagship projects under this initiative. In 2018 the Genome-Phenome Analysis Platform became the fourth resource created by
or with the contribution of RD-Connect partners to be endorsed by IRDiRC with the IRDiRC-Recognized label. The International Charter of Principles for sharing bio-specimens and data received the label in 2015, and the RD-Connect guidelines for the informed consent process in 2016. The FAIR Guiding Principles, which were developed with the involvement of RD-Connect partners from the data linkage team, received the label in 2017.

In conclusion, RD-Connect has achieved and indeed exceeded its goals as a flagship FP7-funded project. It has developed a unique set of infrastructure resources that will go forward into new projects in the future and brought together a diverse community with a collaborative ethos and commitment to data sharing and reuse. Partners can be justly proud of their achievements, and we must also recognise the commitment of the service users, rare disease patients and academic colleagues around the world who have contributed to this success.

Impact on the rare disease research environment

The wide use of the RD-Connect infrastructure is a key indicator for its impact on rare disease research, diagnosis and healthcare. In six years, RD-Connect has become an important player in the rare disease field. The RD-Connect infrastructure has gained visibility and is already being used by many different projects and stakeholders. Important milestones in the use of the infrastructure were the successful collaborations with the major data submitters: NeurOmics, EURenOmics, EuroBioBank, 17 BBMRI-LPC projects, Solve-RD and the European Reference Networks. By October 2018, the GPAP had received 4000 genomic and phenotypic datasets, the Registry & Biobank Finder had recruited 360 rare disease patient registries and 22 biobanks, and the Sample Catalogue had received the data of over 250,000 rare disease biosamples. Through these collaborations, the project has jointly contributed to the discovery of over 100 novel disease genes. This number is expected to increase in the near future since RD-Connect serves as the primary analysis tool for the 17 BBMRI-LPC projects and is also being used by Solve-RD to diagnose unsolved cases from the European Reference Networks.

Each specific work stream had many impacts on the RD research environment.

User-friendly diagnostics and gene discovery

The GPAP has contributed to the RD research field by providing means for broader data sharing and analysis, allowing the identification of new causative variants and genes. This has a clear impact on patients which have received a diagnosis that had eluded previous analysis. Furthermore, it enhances overall RD knowledge, facilitating the diagnosis of future patients and enhancing the understanding of these diseases at the molecular level, thus opening new doors for the development of potential treatments. At the time of writing, 586 users representing 207 organizations are registered on the Genome-Phenome Analysis Platform.

As envisaged when the analysis system was initially set up, the GPAP is thus fulfilling the dual role of enabling data sharing while also lowering the barriers for rare disease researchers to analyse the data they submit themselves even without bioinformatics expertise. Reluctance on the part of researchers to share data due to concerns about losing ownership and being scooped in gene discovery publications was an early hurdle that RD-Connect needed to overcome. The recognition that a mechanism needed to be developed to allow the submitting clinicians and researchers to retain control over their data by empowering them to analyse it themselves was a major success. As has already been shown by the success of the European BBMRI-LPC project (900 exomes) and others such as Consequitur (UK-Turkey collaboration, 500 exomes) which submit data to the system, clinical academics enthusiastically use the system to analyse the patients they themselves have submitted, and this has brought tangible results including gene discovery publications and diagnoses for patients who were previously undiagnosed. The inclusion of even larger amounts of data from new projects such as Solve-RD will further increase this impact.

A catalogue to give visibility to patient registries

The Registry & Biobank Finder is a unique catalogue-style resource that enables the owners of registries and biobanks to publish information about their resource. This increases the visibility of these important resources and provides a mechanism for researchers, particularly those from European Reference Networks, to locate the data and samples they require. This will have an impact on the ability of researchers to do their research, lowering the barriers to locating relevant data.

Quality standards and data linkage plan for registries

A further important achievement was the development of recommendations for registry quality. The recommendations developed by the expert working group are to be used as a framework for improving the quality of RD registries, and this will have an impact on the standards of registries in Europe.
Even more crucially, the strong focus on Findable, Accessible, Interoperable and Reusable (FAIR) data in the registries community is having a substantial impact on the RD field as a whole. It is increasingly recognised that these mechanisms are essential to develop further in order to maximise the value of already generated data, and this is a major aspect of new projects such as the EJP-RD.

There remains a need for coordination between ongoing registry-related initiatives at national and international levels. At national level, we should recommend the development of centralised, public, national, “registry-as-a-service” platforms; centrally resourced platforms will guarantee the access to highly trained staff on the quality of the registry, foster the standardisation of the registries, allow economy of costs and time for setting up new registries, allow to interlink key data sources on different diseases, increase the capacity to develop cooperation at EU level. At a European level, the national platforms for RD registries are collaborating together for creating a centralised European Union-wide framework on patient registries with the benefits of data sharing, reducing duplication of efforts and costs, facilitating validation of results, enabling engagement with experts and the patient community and overcoming the “rare disease problem” in terms of cohort size, powering trials and finding confirmatory cases.

A working community of biobanks

Rare disease biobanks are crucial infrastructures for diagnosis and research. The availability of biospecimens and associated data of patients affected by RD plays a pivotal role in the identification of disease genes and molecular biomarkers, and in the development of novel treatments. RD-Connect developed of a comprehensive platform of tools for making human RD biomaterials and their linked data accessible and available to the scientific community. The Sample Catalogue and Finder tools enabled RD biobanks to join forces to share resources and become an integral part of research infrastructure for RD. We have demonstrated real partnerships between researchers, biobanks, sequencing services, bioinformaticians, analyses and data sharing platform via the BBMRI-LPC research call, where a successful working ecosystem of infrastructures and experts were in place to promote RD research.

Patient engagement in biobanking

Biobanks are encouraged to establish collaborations for collections and dissemination of biological samples. Collaboration in research Calls may be an effective way to ensure biobanking services are available to researchers and clinicians in awarded projects. Similarly, formal collaboration with patient organisations allows precious samples to be collected more systematically from patients and family members, creating critical mass in sample collections for research. The successful collaboration model of TNGB with patient organisations for sample collection especially engaged patients as a key stakeholder in RD research. Another example of patient engagement in WP3 was the representation of an expert patient representative in the Panel for Biobank Assessment. The patient representative can provide advice, stimulate discussions on biobank operations, and facilitate understanding of mutual goals. Such examples have a high impact in stimulating dialogue between biobanks with the patient community as well as fostering their direct engagement in research.

Standards and culture for sharing data and samples from biobanks

A streamlined biobank operation process, spanning from the collection and storage of high quality RD samples/clinical data to their wide distribution, can maximise the impact of biobanks for RD research. RD-Connect designed and facilitated the appropriate usage of RD biomaterials with a recommended workflow that has implications on best practices in sample and data sharing. In addition, sharing of metadata on precious sample collections to a centralised catalogue was reinforced within RD-Connect and EuroBioBank, which contributes to the culture of sharing and FAIR (Findable, Accessible, Interoperable, Reusable) principles. These standards and guides for data/sample sharing are valuable in ensuring biobanks and researchers can work efficiently together as well as building trust during the exchanges. Similarly, the ethical, legal, social implications and challenges on data/sample sharing are applicable to all biobanks. The contribution to the drafting of GDPR Code of Conduct for Health Research has vast implications on the definition a widely recognised code on how to share information in biomedical research. The results and experience of RD-Connect are not only valid for RD research or biobanks, but are valid for other biobanks and research areas.

Innovative bioinformatics solutions

The innovative bioinformatics tools developed as part of RD-Connect will contribute to the achievement of the new IRDiRC goals: 1) All patients coming to medical attention with a suspected rare disease will be diagnosed within one year if their disorder is known in the medical literature; all currently undiagnosable individuals will enter a globally coordinated diagnostic
and research pipeline; 2) 1000 new therapies for rare diseases will be approved, the majority of which will focus on diseases without approved options.

The tools (UMD-Predictor, Human Splicing Finder, VarAFT, ALFA, YABI) to handle sequencing data (genes panels, WES or WGS) can aid in pinpointing disease-causing mutations that can then be experimentally confirmed. These systems can be linked to any existing bioinformatics pipeline and have been included in the RD-Connect Genome-Phenome Analysis platform for research purposes. Clearly, this had a strong socio-economic impact as it significantly contributes to the identification of new disease-causing genes and to reduction of the diagnostic odyssey. In addition, for the most complex situations where the disease-causing gene is unknown, RD-Connect partners released new guidelines and systems (LWAS, WGCNA, YABI, MASTR-MS) to be able to combine multi-omics data (genomics, proteomics, transcriptomics, metabolomics) into a single analysis. This approach also demonstrated its efficiency in various situations and will strongly contribute to the identification of new disease-causing genes. These new systems will strongly benefit to the general public as about 50% of RD genes remains unknown and are responsible for diagnostic wandering.

Because data are numerous but still scattered and difficult to manipulate, WP4 partners strongly supported the FAIR initiative to ensure proper data definition and interoperability, generation of semantic fingerprints for reference -omics and clinical phenotypes. These concepts were successfully translated into reality as illustrated by the SCALEUS (Semantic Web Services Integration for Biomedical Applications) system. While the impact of such initiative might not be directly perceived by the general public, it is a key element for the future to avoid silos and ensure efficient international collaborations to speed-up research.

Clinical diagnosis was aided through an innovative tool for 3D facial analysis. The CliniFace system aims to answer two key research questions: Is it possible to accurately classify syndromes and Human Phenotype Ontology terms from 3D scans of a patient’s face? Are there inferential associations between a patient’s genetic information and their facial (dys)morphology (abnormal features)?

Selection of appropriate treatment options was aided through two complementary approaches: The selection of the best therapeutic approach based on the genetic profile of the patients; and the creation of new therapeutic molecules. The ePGA system is a ‘one stop shop’ Web-based platform to ease the processing, assimilation and sharing of PGx knowledge, and facilitate the aggregation of different PGx stakeholders’ perspectives. The platform offers personalized diagnostics based on reliable genomic/genetic evidence and reduces healthcare costs by increasing drug efficacy and minimizing adverse drug reactions. The second aspect allowed the creation of innovative systems (SKIP-e, NR-Analyzer and Crawfish) that provide genome-wide information to assist researchers and drug companies to efficiently design new molecules to induce exon-skipping (SKIP-e), nonsense readthrough (NR-Analyzer) of trans-splicing (Crawfish), which are three of the most promising approaches for new RD therapies.

Together these "therapeutic systems" will benefit the general population as they will provide easy access to efficient drug selection based on genetic knowledge, on one hand, and, on the other hand potentially facilitate the design of new drugs for RD patients.

ELSI impact and impact of patient involvement

The ethical and legal documents produced in RD-Connect represent the joint effort of the stakeholders involved in the project and are an example of co-produced regulation. This experience demonstrated the importance of the involvement and engagement of patients into the development of policies to achieve standards that reflects common values.

The ELSI leaders engaged in dissemination activities that include more than 30 presentations in scientific conferences, more than 15 publications in scientific journals and organized several public events in which topics were discussed with stakeholders. This has had a substantial impact on the ELSI discussion in the RD community and beyond. RD-Connect also participates in the BBMRI-ERIC Code of Conduct writing group. This group is developing a code of conduct for the implementation of the General Data Protection Regulation (GDPR), which will have a major impact on the way it is implemented for biomedical research. We are ensuring that the patient is at the centre of this development and the Code is directed to scientists working with data in Europe.

The patient-focused activities within RD-Connect have had an impact on the wider patient community, who have been kept informed through dedicated dissemination activities. By presenting at international, European and national workshops, EURORDIS has ensured that all stakeholders interested in the field of RD research were kept abreast of the latest
development of RD-Connect activities and the crucial role that patients played in these activities. Enthusiasm and motivation to participate in the Patient Advisory Council has been ensured by recruiting several new members to join and provide a continuous flow of diverse patient perspectives and this has reinvigorated the group to participate and discuss ethical issues. Frequent and regular communication with the different scientific partners has enabled development of collaborations and increased patient input in activities within RD-Connect. Participation of PAC “delegates” within the “technical” work enabled a direct interaction with the aim to reflect a more active role of the PAC within the project and helps to maintain patient engagement, increase transparency as well as improve patient representatives’ input further within the project and disseminate the projects' activities to the wider patient community. For example, the substantial contribution of the PAC to the upcoming publication “Recommendations for improving the quality of rare disease registries” has demonstrated the added-value of patient engagement in academic publications by bringing much-needed expertise and experience on the topic. Capacity building to improve the knowledge levels of PAC members and the wider RD patient community via the organisation of workshops and webinars on ELSI and scientific issues related to the use of novel genomic technologies, data sharing, the Genome-Phenome Analysis Platform, data protection and the new General Data Protection Regulation have not only increased the impact of PAC members into the project’s outputs but also engaged a high number of patient organisations into the work and ethos of RD-Connect.

Furthermore, linking the activities of the PAC of RD-Connect with that of the patient-centred work in Solve-RD, with the European Advocacy Groups (ePAGs) of the European Reference Networks (ERNs) and with the upcoming EJP-RD will ensure de-duplication of efforts, increase dissemination of the outputs of the projects, improve capacity building for patient representatives, and ensure the longevity of the patient-centred work through its continuation in other projects.

Publications and dissemination activities
To reach a diverse audience, RD-Connect used a range of different dissemination strategies, such as scientific publications, presentations at international conferences, training activities, posters, flyers, articles in the press and online media. In total, RD-Connect has run over 1100 dissemination activities, recorded regularly and reported in the in the deliverable reports. Many of these activities were also announced to the rare disease community through social media and the RD-Connect website. At the end of the sixth year of the project, there are around 490 publications related to RD-Connect, out of which 210 acknowledge RD-Connect and are listed on the project website. The number of publications acknowledging RD-Connect has been increasing each year (Fig 1). A list of RD-Connect peer-reviewed publications is displayed on the website https://rd-connect.eu/scientific-publications/. New publications are highlighted in the newsletter and social media. The major publications on RD-Connect resources include:

To reach a broader audience, RD-Connect has also engaged with the press. Several magazines and journals with different readership and in different countries have published interviews and articles about RD-Connect, including the Rare Revolution Magazine (https://goo.gl/Xxf1k8) Pan European Networks, Horizon (http://bit.ly/2JRs8QL) and more. The value of RD-Connect has been recognised by the European Commission’s Directorate-General for Research, which has promoted RD-Connect as a success story on their website (https://goo.gl/Ymi9uS).

To ensure dissemination of project outputs to the scientific community, RD-Connect was presented in the form of posters, talks, workshops, booths and flyers at numerous national and international events focusing on rare diseases, specific diseases, genomics, and biobanking, including major conferences such as the European Society for Human Genetics (ESHG), American Society for Human Genetics (ASHG), and European Conference on Rare Diseases & Orphan Products (ECRD) and IRDiRC.
Conference. In addition, the work of RD-Connect was highlighted thanks to several prizes awarded to RD-Connect partners, including the Black Pearl Scientific Award for Lucia Monaco (2017) and Volunteer Awards for Chris Sotirelis (2018), JRC Malta Young Scientist Award for Joanna Vella and the ECRD Poster Prize for Dorota Badowska (2018). These activities raised interest in RD-Connect within the rare disease research community, which resulted in submission of large numbers of datasets submitted to the Genome-Phenome Analysis Platform, Registry & Biobank Finder and Sample Catalogue.

Fig. 1. RD-Connect scientific publications and dissemination activities.
RD-Connect organised six annual meetings: Sitges (2013), Heidelberg (2014), Palma (2015), Barcelona (2016), Berlin (2017) and Athens (2018), attended by over 100 people each year. The meetings brought together project partners as well as invited attendees from partner research projects and infrastructures, such as NeurOmics, EURenOmics, EuroBioBank, E-Rare, BBMRI-ERIC, Elixir and the European Reference Networks. The annual meetings provided opportunities to organise training sessions and workshops for external stakeholders and enhance the communication between partners from different work packages. The meeting in Berlin in May 2017 was held back-to-back with the final meetings of NeurOmics and EURenOmics. On 3 May 2017, the three projects hosted a joint open Outreach Day, attended by around 300 participants, including project partners, researchers, patient representatives, policy makers and industry representatives. In three sessions, the participants discussed key cross-cutting topics in rare disease research: data sharing, diagnostics and therapy. The event also included training sessions on the RD-Connect Genome-Phenome Analysis Platform and the Sample Catalogue.

Presence in online media
In its first year of operation, RD-Connect launched the official project website (rd-connect.eu) which has been a major channel for disseminating the project outputs and gave the project a clear identity. It contains information about the project and its outputs as well news (rdconnect.eu/news/) events (rd-connect.eu/events/training/) and downloadable training and dissemination materials, such as webinars, flyers, presentations, manuals, scientific publications and other relevant information. Every year, the website has attracted increasing numbers of visitors. In total, it has received over 120,000 visits from most countries around the world, with highest interest in the UK, USA, France, Italy and Spain. In September 2017, we re-launched the website with a new design and more user-friendly interface. Following that, we saw in average that users displayed more pages per session than before, suggesting increased in the content (Fig. 2). To enhance the engagement with the patient communities, we added the section “For patients and families” (rd-connect.eu/for-patients-and-families/) containing a video introduction and several articles by PAC members, explaining different aspects of rare disease research and the RD-Connect work. The website also contains a several dissemination materials, such as flyers for different stakeholders and presentations.

Fig. 2. RD-Connect website traffic.
After the re-launch of the website with a new, more user-friendly design, the number of page views doubled.
In 2013, the website was supplemented with the RD-Connect monthly newsletter (rd-connect.eu/newsletters/) which contained updates about RD-Connect activities as well as other news from the rare disease field. In total, we have published 46 issues of the newsletter. The interest in the newsletter has been increasing over years, with the highest number of 1458 subscribers in April 2018 (Fig. 3), which dropped after introducing the General Data Protection Regulation in May 2018. Despite of the reduced number of subscribers, we believe the mailing list is targeting the right audience, as since the reduction, the percentage of newsletter opens doubled.
To stimulate communication with stakeholders, particularly with research projects, infrastructures, networks and patient organisations, RD-Connect has been active on social media: Twitter (@ConnectRD, since 2015), YouTube (RD-Connect, since 2016) and Facebook (rdconnect, since 2016). Over the course of the project, hundreds of users have followed these communication channels, with the numbers increasing every month, including individuals as well as organisations, e.g. several ERNs, companies and patient organisations.
The visibility of RD-Connect in the online media has been significantly increased by the launch of the RD-Connect YouTube channel (https://www.youtube.com/channel/UCwvcUPj2fyWGaW13Lva07Aq). A major impact on the popularity of the channel had the RD-Connect explanatory video (https://youtu.be/i0C03vpGhDM) which was released in May 2017 in 7 language
versions and participated in the European Commission’s FP7 showcase. Other videos in the channel include webinars, video tutorials, interviews with diverse stakeholders. We have recorded thousands of views, with the highest numbers from the UK, Italy, Spain, Germany and France.

Fig. 3. Visibility of RD-Connect in online media. The interest in the RD-Connect newsletter, Twitter, Facebook and YouTube channels has been increasing across the duration of the project. The increase in the YouTube views was caused by the release of the RD-Connect promotional video in May 2017. The sharp drop of the number of the subscribers after April 2018 was related to the General Data Protection Regulation, which entered into force in May 2018. The emails of the subscribers who had not confirmed their subscription were erased from the system.

Collaborations
RD-Connect originally worked closely with two EU-funded research projects NeurOmics and EURenOmics, which applied and developed innovative omics approaches to advance research on rare neuromuscular, neurodegenerative and kidney diseases. The end-users from these projects uploaded the first datasets to the RD-Connect systems and provided valuable feedback that helped to tailor RD-Connect to the needs of rare disease researchers. RD-Connect partners have also supported NeurOmics and EURenOmics in data linkage and gene discovery and provided numerous training sessions at their annual meetings. In 2017, the three projects hosted a large joint final conference in Berlin to highlight the research successes of all three projects, and the joint Outreach Day, which provided an opportunity for an open, multi-stakeholder discussion on the key aspects of rare disease research.

After close collaboration since the start of the project, in year 4, RD-Connect established a formal partnership with EuroBioBank, which agreed to be the de facto biobank in RD-Connect. EuroBioBank members, in addition to RD-Connect partners and patient representatives, are members of the RD-Connect Panel for Biobank Assessment.

Throughout the project, RD-Connect interacted closely with the IRDiRC, with RD-Connect coordinator Hanns Lochmüller chairing the IRDiRC Interdisciplinary Scientific Committee and a number of RD-Connect partners contributing to all IRDiRC committees and engaging with various task forces, including the joint IRDiRC-GA4GH task force on privacy-preserving record linkage, which aims to enable the linking of datasets on the same individual across different databases without revealing the individual’s identity. RD-Connect has contributed to the IRDiRC goals of 200 new therapies and the means to diagnose most rare diseases by the year 2020, the first of which has already been achieved ahead of time, in early 2017. Currently, the project is participating in the efforts to achieve the new, more ambitious IRDiRC goals 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, and to develop 1000 new therapies for rare diseases by 2020 (for details see http://www.irdirc.org/irdirc-goals-2017-2027-new-rare-disease-research-goals-for-the-next-decade/).

Linking up with related initiatives is crucial to the project’s impact, success and sustainability. In 2014, RD-Connect established key external collaborations with the biomedical research infrastructures on the European Strategy Forum on Research Infrastructures (ESFRI) roadmap, in particular ELIXIR, BBMRI-ERIC. As multinational research consortia with unlimited duration and a legal entity status in the EU, they are important partners to support future sustainability of RD-Connect and a valuable source of expertise in areas of relevance to RD-Connect. In 2016, BBMRI-ERIC became a full partner of RD-Connect, responsible for assessing sustainability options for RD-Connect’s biobanking resources. Since a large number of RD-Connect partners are also belong to ELIXIR, many activities relating to the rare disease use case, the interoperability and training were carried out jointly, which has been valuable for both sides. This included the joint work on linking registry and biobanks data and co-hosting the “Bring Your Own Data” workshops for researchers working in this area. In summer 2018, BBMRI-ERIC and ELIXIR became members of the RD-Connect Community and joined its Executive Committee.

RD-Connect partners were involved in several Horizon 2020 projects including EXCELERATE (ELIXIR), ADOPT-BBMRI-ERIC (BBMRI-ERIC) and CORBEL (multiple Ris), which led to extensive interactions with Ris, especially in areas relating to data interoperability and mutually beneficial exchange of information and expertise. RD-Connect participates in the Global Alliance for Genomics and Health (GA4GH), particularly in the Clinical Genomics Working Group and the MatchMaker Exchange project, an initiative to encourage secure sharing of information on genetic variants without compromising patients’ privacy. Integration of MatchMaker Exchange allows users learn whether other databases contain their variant of interest or genes with specific variants and contact the data owners to request more information.
To align its outputs with public health infrastructures, RD-Connect has actively engaged with participants in the EU public health sphere, such as the Commission Expert Group on Rare Diseases, the EXPAND project for health data interoperability, RD-Action, the key EU project involved in the development of the European Reference Networks (ERNs).

The inauguration of European Reference Networks (ERNs) in March 2017, international networks of centres of expertise for specific rare diseases, created an opportunity for RD-Connect to increase its engagement in the public health sphere. While ERNs have primarily a healthcare focus, they must also establish research goals. To support the ERNs in the research and diagnostic aspect of their work, RD-Connect provides its expertise regarding sharing research-related clinical, biosample and omics data joined the Task Force on Interoperable data-sharing within the framework of the operations of ERNs and will. In 2018, as part of the Solve-RD project, RD-Connect has started a pilot on re-analysis of existing exomes in four ERNs: GENTURIS, EURO-NMD, ITHAKA, and RND, Six other ERNs -ERK-NET, EYE, GENTURIS, ITHAKA, RITA and ENDO – are interested in participating in re-analysis after the end of the pilot. RD-Connect liaises with all 24 ERNs to host their research sequencing data for analysis and we expect that in longer term all ERNs will become its users.

Global impact
Rare diseases affect around 400 million people worldwide and can be a particular burden for those living in less favoured areas and low-resourced environments, where the means to provide diagnosis and treatment are often scarce. With the decreasing sequencing costs, the technique becomes affordable in more and more countries, and therefore the demand for training in data analysis and interpretation is increasing around the world. As RD-Connect is available to researchers and clinicians around the world free of charge, it may have a particularly big impact in low-resourced regions, where it is already used for clinical diagnosis outside the pure research setting. To target the key regions, RD-Connect established various collaborations and participated in regional and national conferences in numerous countries, including Central and Eastern Europe, Middle East, East Asia and Latin America.

Since 2014, the active engagement of RD-Connect partners in the activities of the global research network Genomic Medicine Alliance and the Golden Helix Foundation led to numerous joint research and educational activities stimulating genomic medicine research, yielding concrete improvements, research projects and partnerships in several countries, such as the collaboration with the SERBORDISInn project (http://serbordisinn.rs/) in Serbia. We particularly need to highlight the rare disease genomics-focused Golden Helix Symposia in Egypt, Malaysia, Serbia and United Arab Emirates, and Golden Helix Summer Schools in Greece.

RD-Connect collaborates with 3Gb-TEST, an EU project aiming to raise awareness of innovations in molecular testing among healthcare professionals to help implement diagnostic genome sequencing in Europe. Both projects have organized together courses on Next Generation Sequencing (NGS) in diagnostic setting in in Prague, Czech Republic (2015); Lisbon, Portugal (2016) and Ljubljana, Slovenia (2017), attended by many researchers and clinicians from the EU priority zones as well as from outside the EU. RD-Connect also collaborates with RD-Action, a project working on data and policies for rare diseases. In 2017, they organised the national Europlan Workshop in Prague, well attended by participants from Central and Eastern Europe.

RD-Connect is also liaising with the Initiative on Rare and Undiagnosed Diseases in Japan, to support supports Japanese researchers in learning how to FAIRify their patient registries data and link them to the data in international data infrastructures. Researchers from the Japan Agency for Medical Research and Development and other institutions several times have visited the RD-Connect coordination office and attended the RD-Connect annual meetings to discuss potential collaboration and sharing of expertise.

To facilitate our engagement with rare disease research communities around the world, we released the RD-Connect promotional video in 7 language versions: English, French, Spanish, Italian, German, Russian and Arabic, and with subtitles available in 48 languages.

The above-mentioned dissemination activities helped to raise the interest in the RD-Connect infrastructure in numerous countries around the world. As a result, the RD-Connect GPAP has currently registered users from many non-Western European countries: Bulgaria, Canada, Croatia, Czech Republic, Egypt, Greece, Hungary, India, Iran, Israel, Latvia, Malta, Romania, Serbia, Slovenia, South Africa, Turkey, Ukraine and USA.

Sustainability
The resources generated by RD-Connect are not time-limited and need to be sustained beyond the initial FP7 funding. Sustainability objectives included consolidating RD-Connect’s position in the community as the leading resource for access to
biosamples and patient registries and for analysis of genome-phenome datasets for diagnosis and gene discovery, expanding collaborations with new data submitters and with European and global initiatives for rare disease research, and ensuring the sustainability of the RD-Connect resources through new sources of funding. We have successfully ensured that the resources developed by RD-Connect will continue to grow and develop thanks to a range of new funding and sustainability mechanisms, in particular the Horizon 2020 project Solve-RD, the upcoming European Joint Programme for Rare Diseases, collaboration with the ESFRI research infrastructures ELIXIR and BBMRI-ERIC, and the launch of the RD-Connect Community.

Beyond the resources themselves, one of the most significant achievements of the RD-Connect project as a whole was the creation of a unique, multidisciplinary community bringing together experts from different fields to work together on enhancing rare disease research and data sharing. To continue this work, the RD-Connect Community was officially launched in July 2018 as an independent, non-governmental, not-for-profit, international association of individuals and organizations sharing the vision of building an open community that works to improve rare disease research. Its mission is to promote, facilitate and accelerate rare disease research by maximizing the availability and (re)use of rare disease data and biosamples through provision of infrastructure, tools and services to share, analyse and link datasets and biosamples in a secure and regulated way. By promoting data sharing and analysis tools and the data sharing ethos among rare disease researchers and clinicians and by raising awareness among patient communities and policy makers, the community aims to maximise the impact of the tools and services developed by RD-Connect members. Interest in the Community is high – in the first three months of its operation, the Community received membership requests from 127 individuals, academic research groups, organisations and European Reference Networks from 29 countries in all continents.

Achieving sustainability through inclusion of the RD-Connect tools in new projects has satisfied the European Commission’s goal of building on existing infrastructure and capitalising on previous investment. This was in fact clearly mandated for one of the major initiatives to be approved recently, the European Joint Programme for Rare Diseases. This major initiative brings together more than 130 entities from 35 countries (including 27 EU Member States, seven Associated Countries and Canada) and has a total budget of over €100 million. It will formally launch in January 2019 and all the major RD-Connect resources play important roles. This includes not only outputs such as the Genome-Phenome Analysis Platform (GPAP), Registry & Biobank Finder, Sample Catalogue and bioinformatics tools, but also the data stewardship expertise and knowledge-transfer activities that grew up within the RD-Connect environment, including the application of the FAIR (Findable, Accessible, Interoperable and Reusable) principles to rare disease data and the many training activities such as the summer school for registries and biobanks and the training for patient representatives facilitated by RD-Connect.

In addition to the EJP-RD, the launch in January 2018 of Solve-RD, a major new Horizon 2020 project in which the RD-Connect platform plays a central role, has been an important new development. This project will see the submission of 19,000 new datasets to the RD-Connect Genome-Phenome Analysis Platform, which will be analysed by researchers from the European Reference Networks for rare neurological disorders, neuromuscular disorders, intellectual disability and genetic tumour risk syndromes. The continued work with the European Reference Networks, now including data submission through Solve-RD, is another highlight: these important international initiatives are now a major feature of the RD landscape in Europe and many of them are looking to RD-Connect to provide not only the infrastructure for omics data sharing and analysis but also the expertise in data stewardship including ontologies and FAIR data principles. In addition, the ESFRI research infrastructures ELIXIR and BBMRI-ERIC have become even more closely associated with RD-Connect, with BBMRI-ERIC committed to working with RD-Connect on the sample catalogue and registry and biobank finder, and ELIXIR incorporating RD-Connect data stewardship and data analysis activities under the banner of its Rare Disease community within the ELIXIR Human Data community.

RD-Connect operates within the context of the International Rare Diseases Research Consortium (IRDiRC) as one of the EU’s flagship projects under this initiative. In 2018 the Genome-Phenome Analysis Platform became the fourth resource created by or with the contribution of RD-Connect partners to be endorsed by IRDiRC with the IRDiRC-Recognized label. The International Charter of Principles for sharing bio-specimens and data received the label in 2015, and the RD-Connect guidelines for the informed consent process in 2016. The FAIR Guiding Principles, which were developed with the involvement of RD-Connect partners from the data linkage team, received the label in 2017.

In conclusion, RD-Connect has achieved and indeed exceeded its goals as a flagship FP7-funded project. It has developed a unique set of infrastructure resources that will go forward into new projects in the future and brought together a diverse
community with a collaborative ethos and commitment to data sharing and reuse. Partners can be justly proud of their achievements, and we must also recognise the commitment of the service users, rare disease patients and academic colleagues around the world who have contributed to this success.

List of Websites:
https://rd-connect.eu/

Reported by
UNIVERSITY OF NEWCASTLE UPON TYNE
United Kingdom

Subjects
Scientific Research

Last updated on 2019-04-11
Retrieved on 2019-07-03

© European Union, 2019