Rare2030

Foresight in Rare Disease Policy

A knowledge-base summary:

RARE DISEASE DATA COLLECTION AND UTILISATION

https://www.rare2030.eu/our-work

TABLE OF CONTENTS

1. INTRODUCTION TO THE TOPIC

The topic of 'data collection and utilisation' is extremely broad. This document therefore contains *select* (i.e. far from exhaustive) summaries of the status quo in a few key areas, including registration, inventorying and coding of diseases, data interoperability, and ethical legal and social issues (ELSI). Other aspects of the topic, for instance those more relevant to diagnostics, will appear in alternative subgroup documents.

Data on any rare condition is extremely precious. No single country will see a sufficient number of patients with any very rare disease to fully understand the condition, in terms of its epidemiology (e.g. how many cases exist in any given population), the range of symptoms observed, the development of the disease over time, and the likely outlook for newly-diagnosed patients. Capturing structured data, based upon field-appropriate standards and ontologies, is particularly important in diagnostics (see Knowledge Base Summary on Diagnostics). Rare disease patient data, especially if collected in a standardised form, takes on greater power to serve what one may loosely term 'secondary purposes', particularly in the case of registry data.

These topics appear in some of the 'foundational' European policy documents in various ways:

CODING AND INVENTORYING:

Commission Communication on Rare Diseases: Europe's challenges (2008) [679 final] Section 3.1. Improving Recognition and Visibility on Rare Diseases:

"To improve diagnosis and care in the field of rare diseases, appropriate identification needs to be accompanied by accurate information, provided and disseminated in inventory and repertory formats adapted to the needs of professionals and of affected persons.[..] The Commission therefore aims to put in place a thorough coding and classification system at European level…"

Council Recommendation of 8 June 2009 on an action in the field of rare diseases (2009/C 151/02. II. ADEQUATE DEFINITION, CODIFICATION AND INVENTORYING OF RARE DISEASES

- Use for the purposes of Community-level policy work a common definition of rare disease as a disease affecting no more than 5 per 10 000 persons.
- Member States (MS) were asked to "ensure that rare diseases are adequately coded and traceable in all health information systems"
- MS were also asked to "Contribute actively to the development of the EU easily accessible and dynamic inventory of rare diseases based on the Orphanet network and other existing networks as referred to in the Commission Communication on rare diseases"

In 2014, the Commission Expert Group on Rare Diseases adopted a Recommendation on Ways to Improve Codification for Rare Diseases in Health Information Systems

In 2017 and 2018, RD-ACTION – the EU Joint Action for Rare Diseases- generated several practical outputs to build upon this Recommendation and support countries in implementing the OrphaCode.

REGISTRIES:

Commission Communication on Rare Diseases: Europe's challenges (2008) [679 final]

Section 5.11. "Registries and databases constitute key instruments to increase knowledge on rare diseases and develop clinical research … A key issue will also be to ensure the long-term sustainability of such systems, rather than having them funded on the basis of inherently precarious project funding."

The Council Recommendation of 2009 asked Member States to "Consider supporting at all appropriate levels, including the Community level, on the one hand, specific disease information networks and, on the other hand, for epidemiological purposes, registries and databases, whilst being aware of an independent governance"

One of the eight sets of Recommendations adopted by the EUCERD and Commission Expert Group for Rare Diseases was dedicated to registration and patient data collection. The EUCERD Recommendations on Rare Disease Patient Registration and Data Collection (2013) remain an important compendium of high-level principles for judicious creation and operation of registries.

NB. Naturally, there is an extensive list of policies, Regulations and Directives with a bearing upon this broad topic which, whilst not RD-specific, should obviously be considered 'core' to this subject; for instance

- the General Data Protection Regulation ((EU) 2016/679) which came into force in May 2018
- the Directive on the Application of Patients' Rights in Cross-border healthcare (Directive 2011/24/EU), from the perspective of data moving across borders
- the 2018 Commission Communication on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society

2. RARE DISEASE REGISTRIES

With Registries have traditionally been viewed as an excellent way to collect and pool patient data. The WHO defines a registry as "a file of documents containing uniform information about individual persons, collected in a systematic and comprehensive way, in order to serve a pre-determined scientific, clinical or policy purpose". Registries collect information on patients afflicted by a particular disease or group of diseases. By combining data on as many patients as possible, at the regional, national, European or global level, the power of the data increases exponentially. Registries, particularly when used by many different centres, enable researchers to accrue a so-called 'critical mass' of patients which would often otherwise be impossible.

2.1 What purposes can Registries serve?

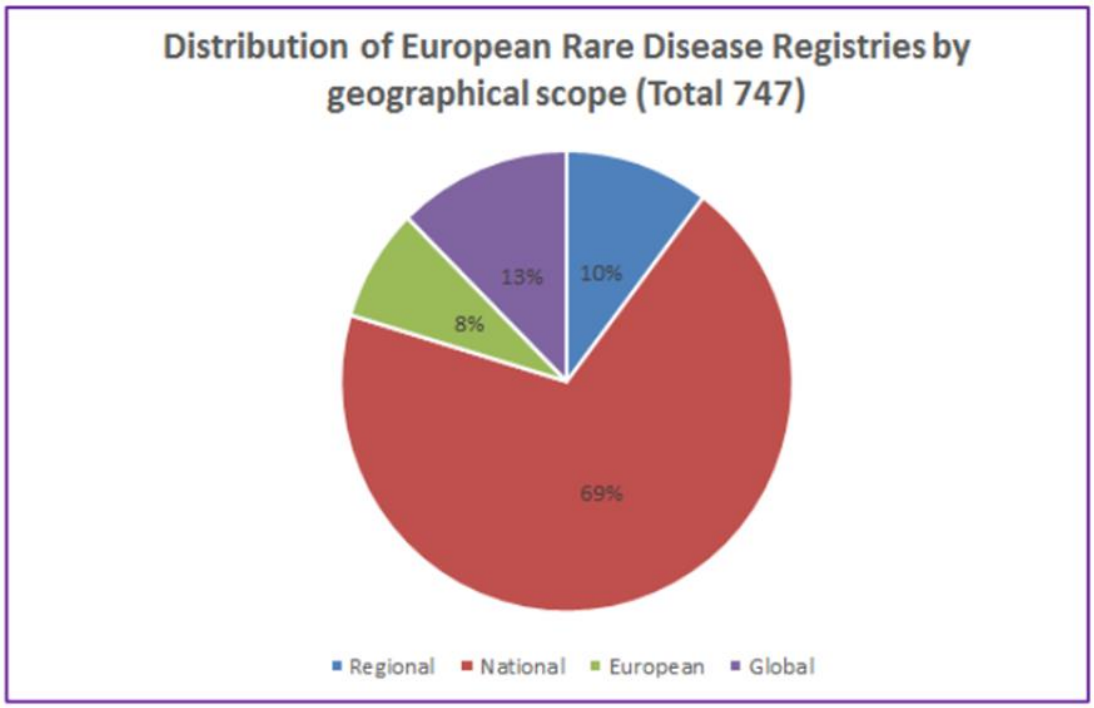
- Registries can focus upon the epidemiology of the disease i.e. how the disease is caused/what are its origins and its impact in any given population (including its rarity). Such epidemiological information is very valuable in assessing disease threats and informing the appropriate planning of health services;
- Registry data can demonstrate the efficacy of different management and therapeutic options, presuming information on treatment regime and clinical outcomes is captured.
- Registries -if established in a certain way - can support the post-marketing surveillance of (conditionally) approved orphan medicinal products
- The correlation between certain genetic mutations and corresponding clinical presentation (phenotype) may be elucidated by registry data. Sometimes patients with the same condition and the same genetic mutation exhibit very different symptoms and experience the disease with varying severity: only by capturing this information routinely and robustly are researchers better able to understand rare conditions and their prognoses by correlating patients' genotypes and phenotypes (in other words, understanding how different combinations of genetic anomalies result in particular clinical presentations).
- Registries are a significant enabler for clinical research, for instance by supporting an assessment of the feasibility of conducting a trail in the first place, and later by facilitating the recruitment of patients. This is particularly useful when registries record an accurate genetic diagnosis (i.e. they stipulate the particular mutation responsible for causing the condition). As medicines and interventions become more personalised, clinical trials often target a specific mutation and therefore need to recruit a particular sub-set of patients. The existence of detailed genotypic information enables a sponsor to assess the number of trial participants they could potentially recruit, and where they are based.

2.2 What is the status quo of rare disease registration in Europe?

Information of the European status quo regarding rare disease registration is available in several fora (with more information likely to emerge through overarching initiatives such as the EU Joint Programme Co-Fund for RD Research, ERN mapping exercises, etc.)

According to the May 2018 Orphanet Report Series report 'Rare Disease Registries in Europe' (2019 update due very soon!) **there are 747 disease registries in Europe: 51 operate at the European level; 93 Global; 518 National and 77 Regional.**

Most of the registries are established in academic institutions. A minority are managed by pharmaceutical or biotech companies, with others being run by patient organisations. A full list, based upon the data contained in the Orphanet database, is available here - http://www.orpha.net/orphacom/cahiers/docs/GB/Registries.pdf.



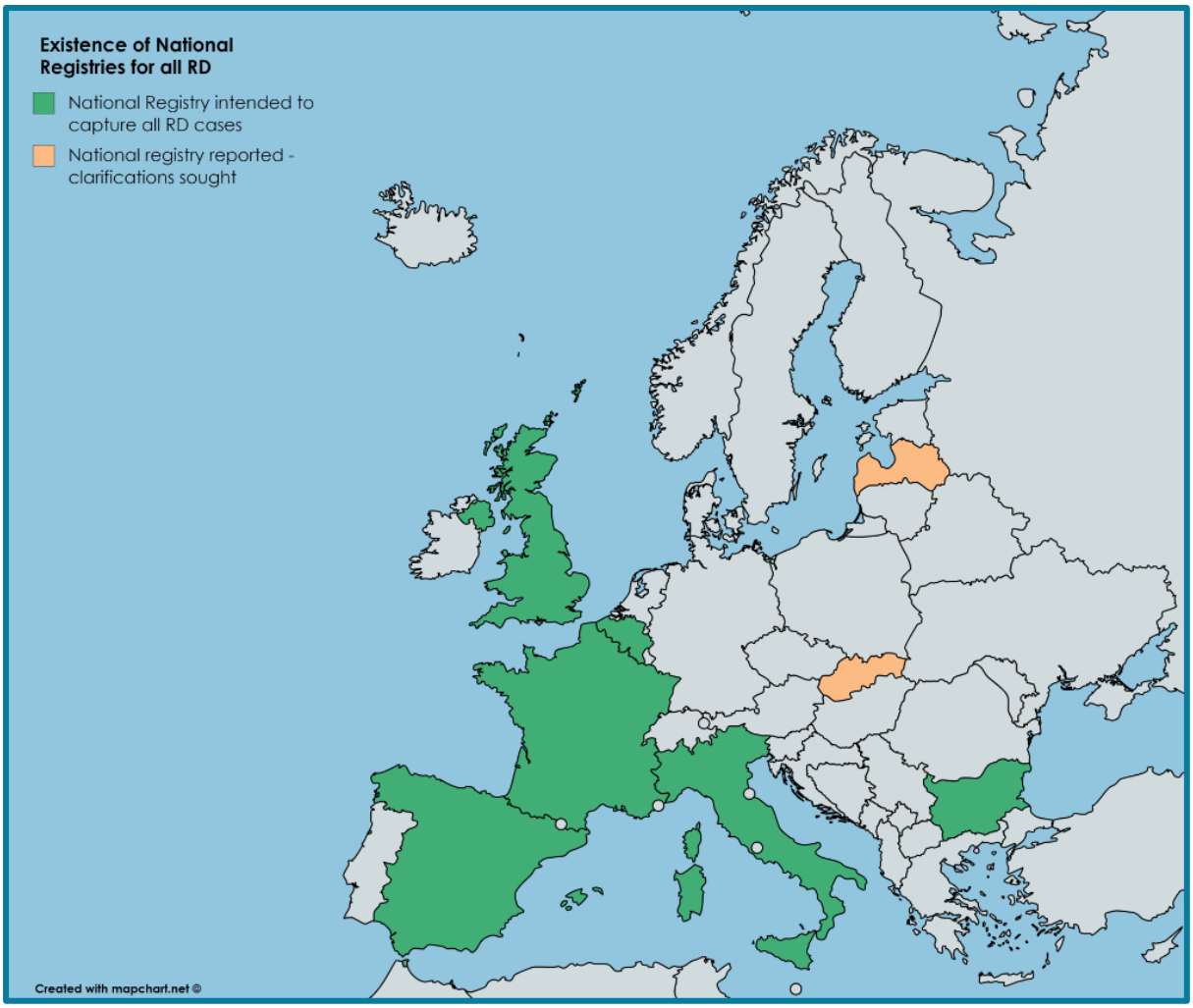
Information on national activities concerning RD registries is also elicited from each EU country via the Resource on the State of the Art of Rare Disease Activities in Europe. According to the latest collection (as of May 2019 - data is still being updated in some countries), there is quite a heterogeneous reality across Europe as regards **national registries** designed to capture all cases of a rare disease in the national territory:

The following countries reported the existence of a **national-level registry established/evolved specifically for RD patient cases (i.e. to register any patient with a RD):**

- **Belgium**: The national level Central Registry for Rare Diseases (CRRD) is prospectively collecting a limited set of variables, having started with a proof-of-concept phase in two genetic centres after which the other six recognized genetic centres came on-board.
- **Bulgaria**: In 2017, a project was established to create a National Register of Patients with Rare Diseases. The registry appears operational as it is already collecting a number of data items including patient's name, date of birth, leading diagnosis, accompanying diagnosis, examinations, studies, consultations, etc. family history, etc.
- **France**: Has the project named BNDMR (Banque Nationale de Données Maladies Rares-National National Rare disease Bank). This was initially intended to develop and accelerate research projects; however, the concept is being further developed and it will be possible to allow mapping of patients' needs and healthcare received, and to facilitate patients' recruitment for clinical and epidemiological studies and clinical trials. BNDMR is populated via two main data streams: BAMARA, which is a care data collection; and DPIs (a DPI is the Patient Medical File each hospital completes)
- **Italy**: Has a national registry for RD, functionally linked to regional and interregional registries of RD. This was established through Art. 3 of the Ministerial Decree of the 18th May 2001 No 279. The National registry is based at the National Institute of Health. Regional/interregional registries are managed by Regional Health authorities.
- **Spain**: In 2011 the Carlos III Institute of Health (ISCIII) joined the International Rare Disease Research Consortium (IRDiRC) and launched an internal and strategic IRDiRC call for Spain, which resulted in the consolidation of the Spanish Registry Network for Research for Rare Disorders (SpainRDR). More

recently, the passage of Royal Decree 1091/2015 created and regulated the *State* Registry of Rare Diseases.

- **UK**: In 2015 the long-standing congenital anomalies registry network evolved into the broader National Congenital Anomaly & Rare Disease Registration Service (NCARDRS), in England. Similar systems to NCARDRS are being considered –or indeed now being implemented- in Scotland, Wales and Northern Ireland. Since 1998, Wales has operated the CARIS (the Congenital Anomaly Register and Information Service) . In 2018, Scotland launched CARDRISS (the Congenital Anomaly and Rare Disease Registration and Information Service) which is now operational (*new data: more details will be gathered)*
- **Slovak Republic**: The national registry for rare diseases was created in January 2014: it is capturing all cases of hereditary diseases, chromosomal anomalies and genetic syndromes (*new data: more details will be gathered*)
- **Latvia**: Since 2015, rare disease registration is implemented under the Register of congenital anomalies, which is apparently broadened to include all RD cases (*new data: more details will be gathered*))



Several other countries reported in their 2019 updates that concrete steps towards a national registry were now underway e.g.

- **Croatia** has begun to collect data for a potential registry and the Croatian Society for Rare Disease and the Croatian Medical Association has funded the creation of the software needed for a national rare disease registry.
- **Hungary** also began development of National RD registry software
- **Malta** is seeking to link all cases of RD appearing in their other existing national registries

Beyond Europe, several countries have established national RD registries, for instance, **Colombia** now has a national registry for rare diseases. In the **USA**, the Office of RD Research launched a pilot project in 2012 to establish the Global Rare Diseases Patient Registry and Data Repository (GRDR). By 2016, the GRDR had agreed Common Data Elements (CDEs) organized into 10 categories that include required and optional elements, and has launched consent forms and information resources. In 2017, the GRDR changed its name to the Rare Disease Registry (RaDaR) Program

2.3 What initiatives are supporting rare disease registration, and in what way?

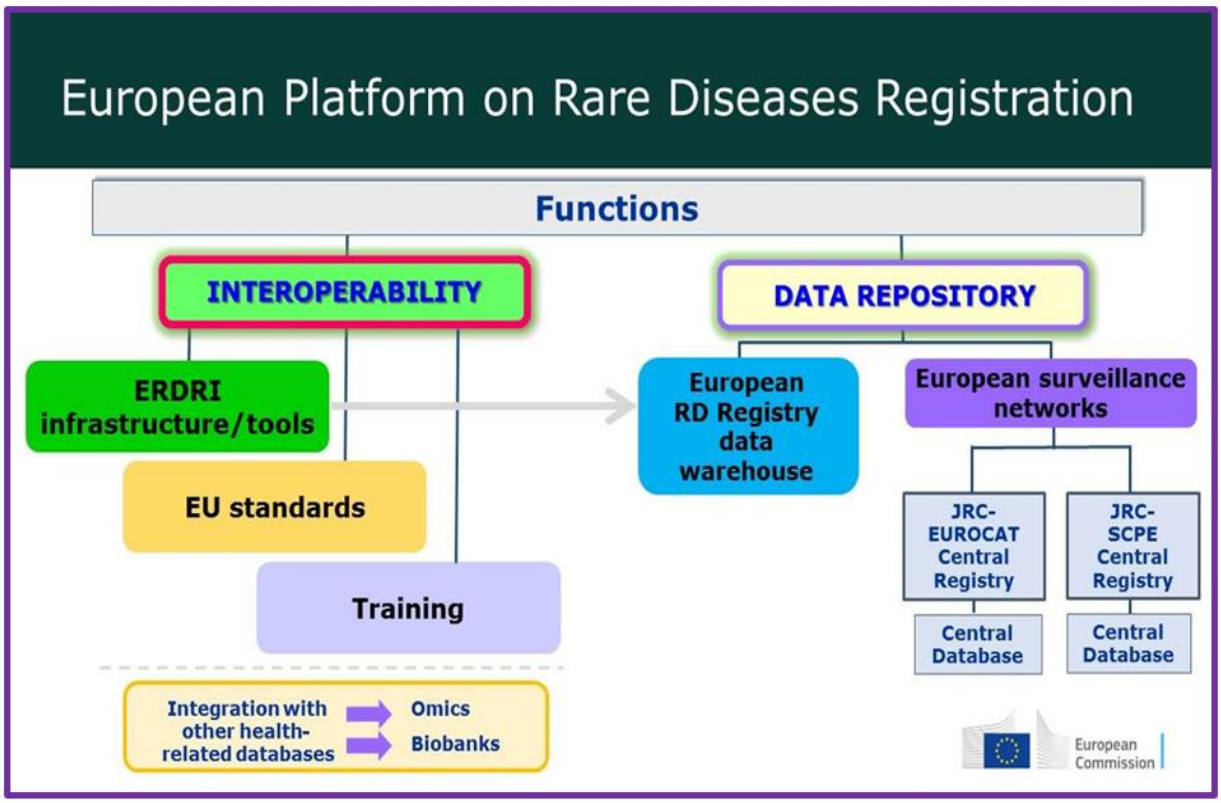
Please note that the following table is selective – for a more exhaustive summary see for instance Overview Report on the State of the Art of Rare Disease Activities in Europe, 2018 Page 65 onwards)

Initiative/Project	Brief Outline	Key Resources/Contribution to the field
EC Joint Research Centre	Signed an Agreement in 2013 to establish a European Platform on RD Registration. Actions are ongoing and are RD-specific Main goal – addressing the lack of interoperability in Europe's RD registries	• Resources to support the various elements of the ERDRI (EU RD Registration Infrastructure), including: • Common Data Set for RD Registries (based on EUCERD Joint Action, RD-Connect, and EPIRARE outputs) • ERDRI User Access Guide • (see further, below)
EMA Patient Registries Initiative	Established in 2015. Actions are ongoing Not RD-specific. Main Goal - facilitating interactions between registry coordinators and potential users of registry data, both at an early stage of therapy development and during the MA evaluation procedure and post-authorisation	• Discussion Paper: Use of patient registries for regulatory purposes(2018) • Inventory of Patient Registries (within the EnCePP Resources Database) • Reports on Qualification of two registry networks and reports from disease-specific workshops here
EJP for Rare Diseases	European Joint Programme on rare Diseases, Pillar 2, has a particular focus on Registries.	It will develop: • **A Centralized metadata repository** describing pre-existing resources (including catalogues, data repositories, tools and infrastructures) with rare disease-specific semantic standards and metadata which conforms to an ontological, machine-readable model. • **A federated ecosystem of FAIR-at-the-source resources**, in order to enable data discovery, sharing and analysis down to the record level

ERN Registry Grants (DG SANTE)	5 ERNs were funded to establish new/link existing registries in their field, back in 2018. A second call was launched for the other 19 in 2019. The main purpose of the 5 funded registries appears to be creating a tool to register all patients visiting the HCPs of which each ERN is composed, collecting well-defined datasets. These registries are building links to other existing disease registries	• Plans and priorities of the 5 funded ERN registry projects are available via their individual websites (you can find these here p51) • The call for registry-support for the other 19 Networks will close in September 2019. Collaboration across ERNs here, in terms of dataset selection and platform sharing, is being encouraged
RD-Connect	FP7 Initiative 2012-2018, establishing a platform to support RD research by linking data from biobanks, registries, databases and bioinformatics. Funding period expired	• Developed Registry ID Cards – designed to improve the accessibility and usability of existing RD registries by providing each with an ID card. Registries were enrolled to the RD-Connect Registry and Biobank Finder
PARENT Joint Action	Cross-border Patient Registries Initiative (PARENT JA) Funded via the 2nd Public Health Programme from May 2012 until November 2015 (funding period expired)	• Developed Methodological Guidelines and Recommendations for Efficient and Rational Governance of Patient Registries, along with several other key outputs. • This output now exists as a Wiki (http://parent-wiki.nijz.si/index.php?title=Methodological_guidelines_and_recommendations_for_efficient_and_rational_governance_of_patient_registries) and was formally endorsed by the eHealth Network (eHN) in 2015

2.4 European Platform for Rare Disease Registration

In December 2013, the European Commission's Joint Research Centre, in collaboration with DG SANTE, initiated development of the European Platform on Rare Diseases Registration (EU RD Platform) to address the serious fragmentation of rare disease patient data contained in hundreds of registries across Europe. The services and tools to be offered by this Platform have become much more clear and concrete in recent years, and a high-level summary is therefore presented below (see further https://eu-rd-platform.jrc.ec.europa.eu/)



(Image courtesy of JRC: as utilised in the Overview Report for the State of the Art of Rare Disease Activities in Europe)

The Platform has two main functions, as above: Interoperability and Data Repository

1. Searchable, queryable and findable RD patient data across RD patient registries (Interoperability)

This achievement, requested for many years by the RD community, is based on the development of the European RD Registry Infrastructure (ERDRI), which contains the following main components:

- the European Directory of Registries (ERDRI.dor) which gives an overview of the RD registries joining the Platform, with their main characteristics and description;

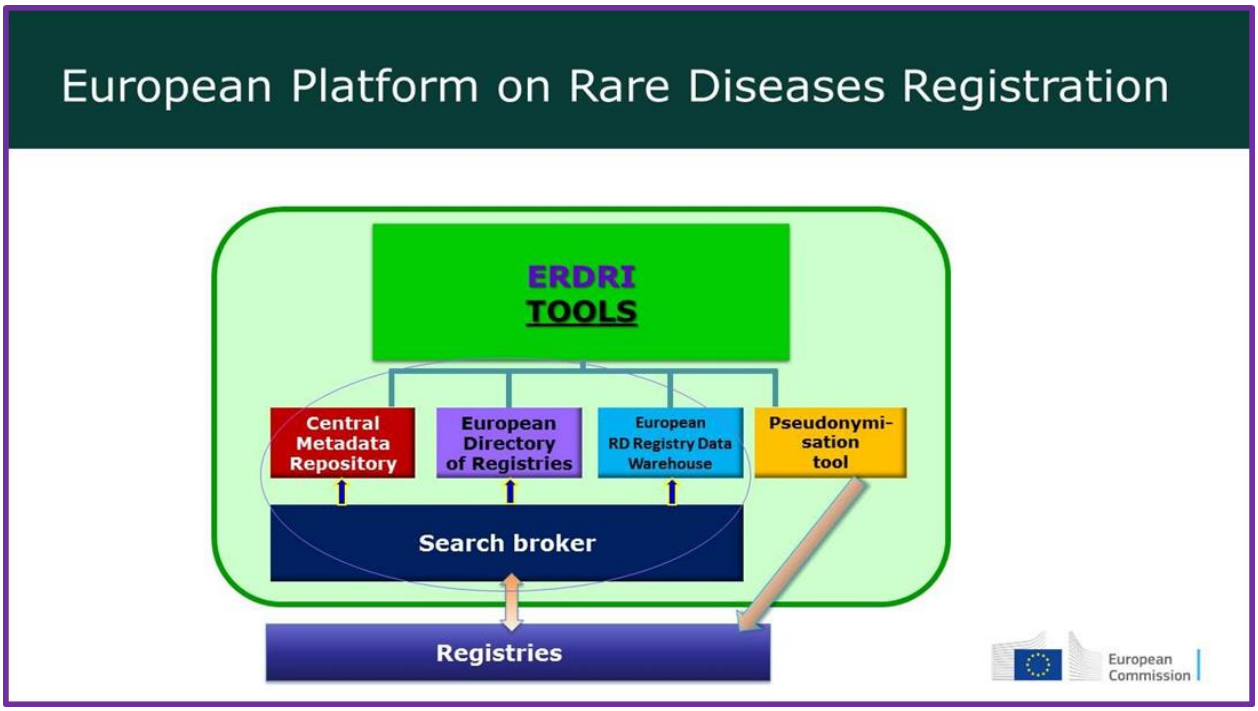
- the Central Metadata Repository (ERDRI.mdr) which ensures semantic interoperability between RD registries;
- the Pseudonymisation Tool (EUPID) providing pseudonyms to participating registries;
- a Search broker helping to retrieve data of interest

The European Commission's JRC also offers training on the tools and functions provided

2. Data Repository

The EU RD Platform provides:

- the European RD Registry Data Warehouse (data repository), which will contain aggregated data from the RD registries. This is facilitated by the promotion of a single set of **common data elements** (see table above)
- the central data repositories (and function of Central Registries) for two long-established surveillance networks: EUROCAT (congenital anomalies) and SCPE (cerebral palsy in children and young people). This activity involves more than 40 registries for EUROCAT and more than 20 for SCPE; therefore, establishing these repositories and central registries was a complex legal and organisational process.



3. EUROPEAN DRIVE TOWARDS INTEROPERABILITY AND REUSE OF RARE DISEASE DATA

Significant emphasis has been placed in recent years -via a number of cross-cutting disease -agnostic projects (e.g. EU Joint Actions for Rare Diseases, RD-Connect) on capturing data about RD patients in a standardised way, to allow some degree of pooling/sharing/querying of that data. An important step forwards, in terms of clarifying the best standards and approaches (e.g identifying the most appropriate ontologies) has been the emergence and greater visibility of the **FAIR data principles**.

The FAIR principles originated outside of the RD field but are especially pertinent in domains which necessitate a significant level of data 'sharing'. FAIR is an acronym, standing for Findable, Accessible, Interoperable, Reusable. The concept was developed by a team of scientists and data experts led by Prof. Barend Mons and has –particularly since publication of a key 2016 paper - gained traction globally: organisations which endorse FAIR data principles include ELIXIR, BBMRI, the European Open Science Cloud, FORCE11, NIH through its 'commons' program, and the G20. The FAIR principles acknowledge that actually *exchanging* data between centres and certainly between jurisdictions is challenging. Instead, 'FAIR' promotes the concept of making data *queryable*, which is an efficient -and far more achievable- goal. A key publication is http://www.nature.com/articles/sdata201618 and there is a useful introduction to using FAIR concepts here.

In 2017, a number of fields established **GO-FAIR Implementation Networks**, designed to unite stakeholders interested in promoting the spread of FAIR principles in their particular domain, working towards an ecosystem of FAIR data services. In 2018 a GO-FAIR Implementation Network for Rare Diseases was established, seeking to anchor together the individual 'FAIRification' efforts in the RD field.

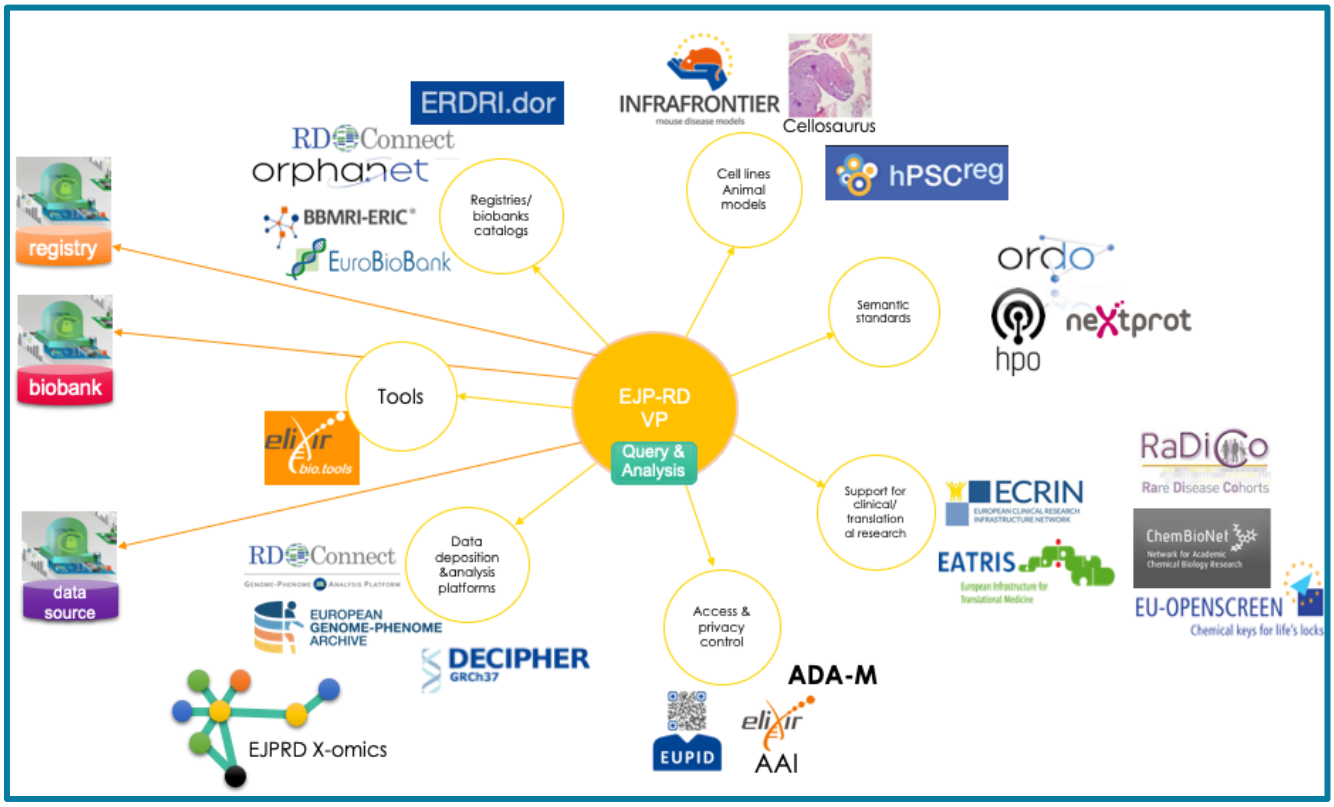
Particular emphasis is placed upon supporting the ERN community to make their data FAIR, given the unique opportunities and economies of scale offered by these new Networks. For instance, the GO-FAIR Network is an opportunity to advance the actions espoused by the 'RD-ACTION Recommended Practices on Standardising Data in the context of the operation of ERNs' relating to FAIR data in the ERN framework.

An important component of making data FAIR is the use of appropriate and agreed ontologies to enhance the visibility of rare disease cases in national health systems and research resources, and to allow the exchange and understanding of such data through (increasingly) electronic formats (see below)

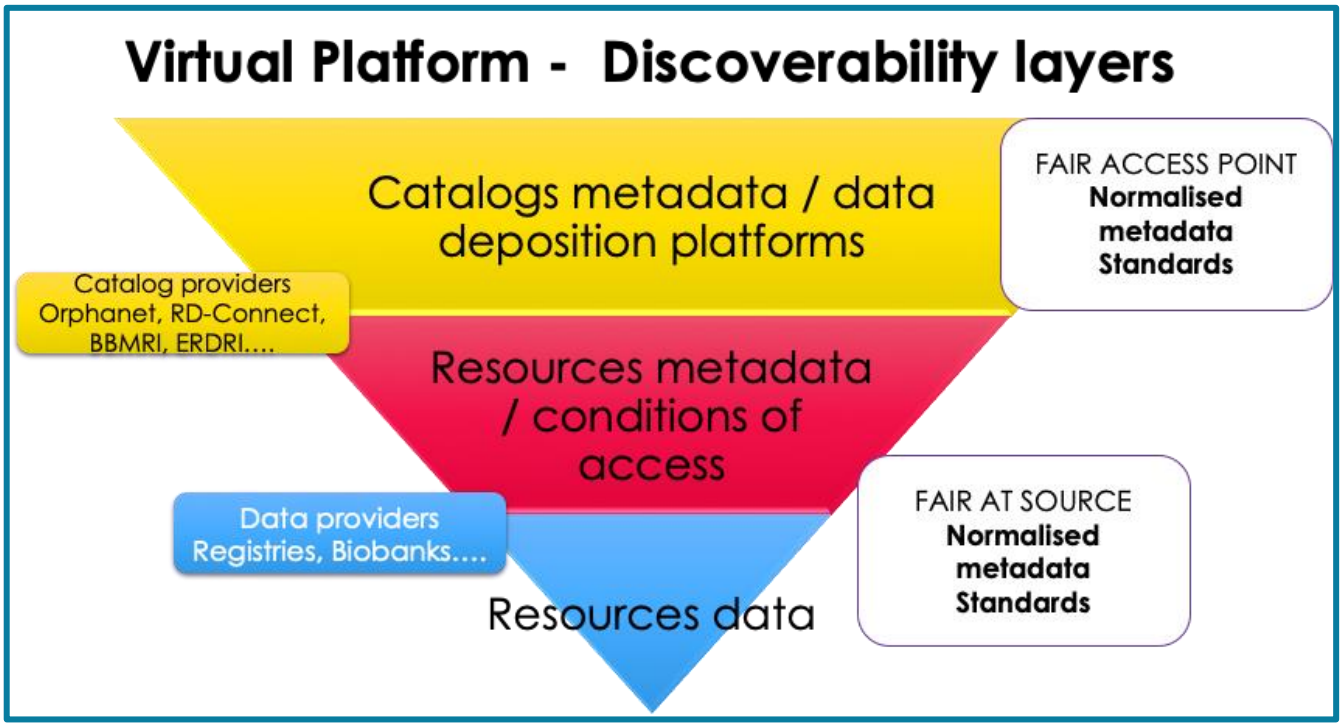
EJP-RD Pillar 2

The European Joint Co-fund Progamme for RD (EJP-RD) will promote and facilitate the implementation of FAIR principles in RD data sources, with a special focus in RD registries. This will be achieved by providing data stewardship support to ERN's registries and providing training on FAIRification.

The main aim of the collaborative work in Pillar 2 is geared towards decreasing fragmentation and maximizing European capacity to enable better and more efficient research on RD by bringing together the interdisciplinary key players, their assets and know-how, to provide coordinated access to resources and data through a common Virtual Platform (VP). These resources either exist already or will be created over time; for instance, RD multi-omics pathways data will be generated and made available, and ERNs registries data will be made discoverable and queryable as these registries are established. The following schema seeks to illustrate the range of resources and actors Pillar 2 of the EJP RD will unite:



The Virtual Platform main concept can be schematised as follows:



In line with the Council Recommendation of 2009 (see above, p1-2), significant progress has been made to increase the visibility of rare diseases in health systems and in research data collections, through use of appropriate nomenclatures and ontologies. Orphanet produces a nomenclature and classification specific for RD http://www.orphadata.org/cgi-bin/rare_free.html, in which each RD has a unique identifier, **the ORPHAcode**. The Orphanet nomenclature is interoperable with other medical terminologies in use (ICD10 and 11, SNOMED-CT, OMIM, MeSH, MedDRA, GARD) and is the backbone of a network of relationships with other data such as genes, phenotypes, functional consequences, epidemiology, related to RD. This network is delivered as an ontology of RD, ORDO.

The ORPHAcode was recently promoted as a best practice by the Commission Steering Group on Promotion of Health and Prevention of chronic non-communicable diseases (SGPP), which resulted in a EU-funded project, **RD-Code** (2019-2021) aiming at implementing the ORPHAcodes in 4 EU countries (Czech Republic, Malta, Spain and Romania) following the guidance and recommendations for codification of rare diseases produced by RD-Action (2015-2018).

Indeed, ORPHAcodes are already being used by the majority of Member States, albeit via diverse implementation models (in centres of expertise, in national registries, in hospitals or in the national codification system). Generalisation of the ORPHAcodes will ultimately allow for improvement of RD patients' visibility and traceability in health systems, and for a better epidemiological knowledge across Europe.

Further to the recognition of a rare disease diagnosis in health systems and registries, a standardised characterization of the clinical *manifestations* (phenotypes) of rare diseases is crucial to improve recognition of conditions by doctors and for RD patient match-making and genomics interpretation. **The Human**

Phenotype Ontology (HPO) is now the standard terminology and ontology for RD phenotyping (and indeed has secured the status of 'IRDiRC-Recognised Resource'). HPO was developed at the Charité (Berlin, Germany) and it is now run by the Jackson Institute (USA). HPO and ORDO are usable together as an ontological ecosystem, HOOM (HPO-ORDO ontology module). This was made possible through an eRARE-funded project, HIPBI-RD.

3.1 Codification of Rare Diseases and Capture of Phenotypic Features

In line with the Council Recommendation of 2009 (see above, p1-2), significant progress has been made to increase the visibility of rare diseases in health systems and in research data collections, through use of appropriate nomenclatures and ontologies. Orphanet produces a nomenclature and classification specific for RD http://www.orphadata.org/cgi-bin/rare_free.html, in which each RD has a unique identifier, **the ORPHAcode**. The Orphanet nomenclature is interoperable with other medical terminologies in use (ICD10 and 11, SNOMED-CT, OMIM, MeSH, MedDRA, GARD) and is the backbone of a network of relationships with other data such as genes, phenotypes, functional consequences, epidemiology, related to RD. This network is delivered as an ontology of RD, ORDO.

The ORPHAcode was recently promoted as a best practice by the Commission Steering Group on Promotion of Health and Prevention of chronic non-communicable diseases (SGPP), which resulted in a EU-funded project, **RD-Code** (2019-2021) aiming at implementing the ORPHAcodes in 4 EU countries (Czech Republic, Malta, Spain and Romania) following the guidance and recommendations for codification of rare diseases produced by RD-Action (2015-2018).

Indeed, ORPHAcodes are already being used by the majority of Member States, albeit via diverse implementation models (in centres of expertise, in national registries, in hospitals or in the national codification system). Generalisation of the ORPHAcodes will ultimately allow for improvement of RD patients' visibility and traceability in health systems, and for a better epidemiological knowledge across Europe.

Further to the recognition of a rare disease diagnosis in health systems and registries, a standardised characterization of the clinical *manifestations* (phenotypes) of rare diseases is crucial to improve recognition of conditions by doctors and for RD patient match-making and genomics interpretation. **The Human Phenotype Ontology (HPO) is now the standard terminology and ontology for RD phenotyping (and indeed has secured the status of 'IRDiRC-Recognised Resource'). HPO was developed at the Charité (Berlin, Germany)** and it is now run by the Jackson Institute (USA). HPO and ORDO are usable together as an ontological ecosystem, HOOM (HPO-ORDO ontology module). This was made possible through an eRARE-funded project, HIPBI-RD.

3.2 Electronic Health Records: the European status quo

As Europe moves increasingly to electronic (as opposed to paper) health records, exciting opportunities await in terms of the potential to link the health records of patients living with a rare disease, resulting in such benefits as

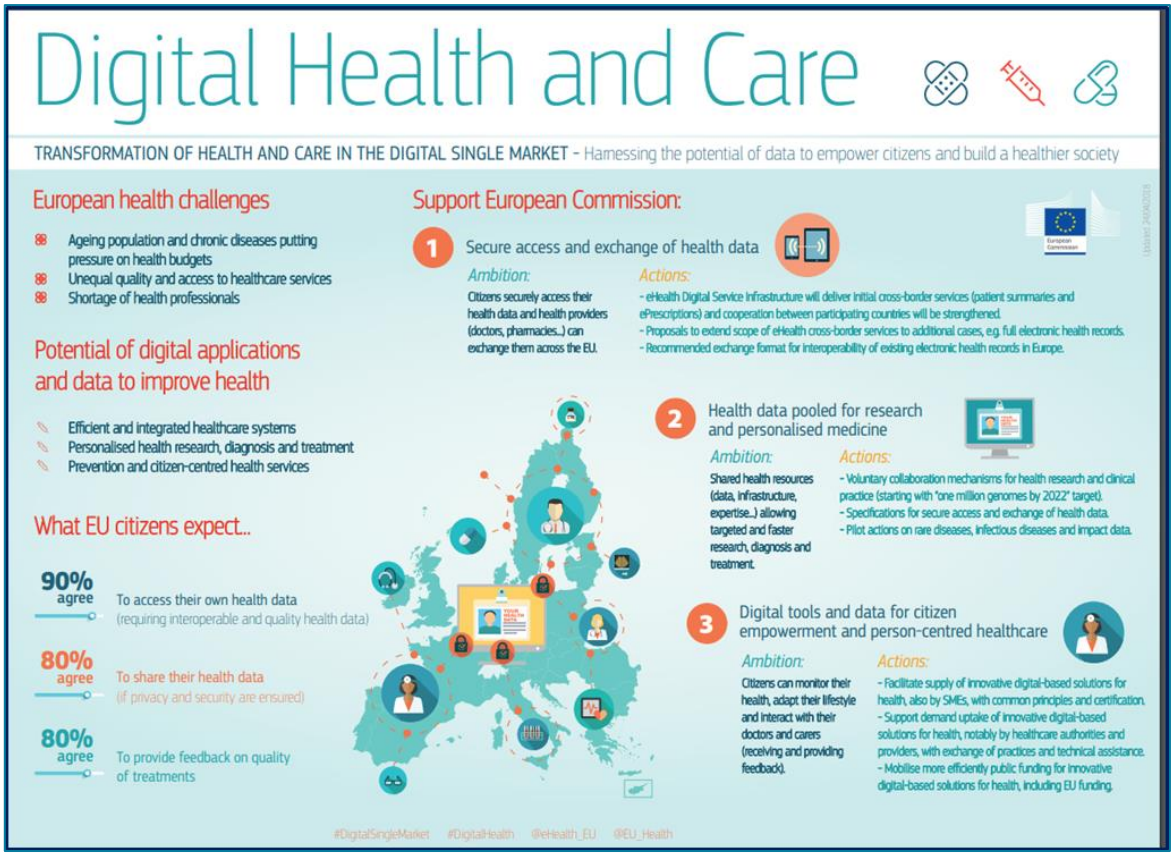
a. reduced need to explain health histories time and again when meeting any new professional; and

b. more streamlined approaches to integrated care, with all relevant encounters (ideally across the health and social spheres) amalgamated to one EHR. A particular benefit, for time-short data entry teams, would be the capacity to populate at least sections of complimentary real-world evidence resources such as registries by automatically extracting relevant data from EHRs. Enriched and well-designed EHRs could also potentially support activities such as feasibility studies and recruitment to clinical trials. Many barriers stand in the way of a seamless integration of EHRs both between geographical jurisdictions (sometimes within) and indeed between EHRs and other complimentary data resources.

Not least amongst these is the fact that European countries are developing their own systems for electronic data capture in the health sphere. An important step to address this fragmentation was the publication in 2018 of the Commission Communication on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society. This document sets out the Commission strategy to transform healthcare under the Digital Single Market, and sets out a number of specific proposals, geared around 3 areas:

1. **Citizens' secure access to their health data, also across borders**- enabling citizens to access their health data across the EU;
2. **Personalised medicine through shared European data infrastructure** - allowing researchers and other professionals to pool resources (data, expertise, computing processing and storage capacities) across the EU;
3. **Citizen empowerment with digital tools for user feedback and person-centred care** - using digital tools to empower people to look after their health, stimulate prevention and enable feedback and interaction between users and healthcare providers.

An important step forwards, in terms of enabling the exchange of health data across borders, is the European Commission drive to **prototype a European interoperable EHT exchange .**



(Infographic taken from https://ec.europa.eu/digital-single-market/en/news/infographic-digital-health-and-care-eu)

3.3 Exchanging data across borders

Once one accepts the need to be able to pool/share/query data held in different national jurisdictions, it is necessary to agree and implement mechanisms (with accompanying legal and social governance frameworks) to enable this. There have been numerous efforts to exchange health-related data across borders: two examples are briefly highlighted below:

***European Reference Networks**:* A key pillar upon which the ERN concept is based is the mantra that wherever possible, data should travel, rather than patients themselves. In reality, this meant the creation of a robust, secure platform to exchange data between HCPs based in different EU MS/EEA countries. The

European Commission supported the provision of a suitable platform, which is today known as the CPMS (Clinical Patient Management System). Before, during, and after the creation of this Platform, efforts were made to ensure data was captured in such a way as to extend the 'life' of that data for secondary purposes, beyond the immediate goal (i.e. the virtual referral to a panel of experts, on diagnostic advice, suitability for specialised procedures, treatment options, etc). The precise nature of these 'secondary purposes' is yet to be determined. For initial discussion, see the RD-ACTION Recommended Practices for Data Standardisation in the Context of the Operations of ERNs.

CPMS in numbers:

- **As of May 2019, 1268 active users are registered in the CPMS (an 'active user' is an individual who has logged in at least once);**
- **623 panels have been opened at some stage**
- **245 panels have been closed and archived.**

An important step in this process was the creation of a common pan-ERN Informed Consent template and information sheet, to authorise the exchange of data for care (and possible additional uses). The Networks are being encouraged to personalise core datasets specific to diseases or groups of diseases addressed by their network, and to implement these datasets with reference to particular ontologies (e.g. the Human Phenotype Ontology or HPO), to increase the interoperability of that data (for a variety of possible future purposes).

eHealth Network: To support the exchange of patient data across borders, the CrossBorder healthcare Directive established (via Art.14) a voluntary body known as the eHealth Network (eHN). The eHN oversaw the creation and evolution of a number of eHealth Digital Service Infrastructures or eHealth DSIs. This work has been funded within the framework of the Digital Europe Programme and can, in some sense, be considered to stem from (or at least was largely driven by) the epSOS initiative. Ending in 2014, epSOS ("Smart Open Services for European Patients") was a European large-scale pilot testing the cross-border sharing of

a) a patient's most important health data summary, intended for use in an unplanned (e.g. emergency) care situation when travelling or working abroad; and

b) b) an electronic prescription (ePrescription).

A small TaskForce initiated under the EU Joint Actions for Rare Diseases has undertaken initial work with eHealth initiatives to highlight the need to consider rare disease patient needs in these two Digital Service Infrastructures. Caring for a person living with a rare disease presents certain specificities that merit the inclusion of additional data elements in the patient summaries to support emergency care or planned cross-border healthcare.

4. ETHICAL DATA MANAGEMENT AND DATA PROTECTION

Collection and use of patient health-related data is, naturally, subject to strict regulations. In Europe, the EU General Data Protection Regulation (Regulation (EU) 2016/679) (GDPR), effective on May 25, 2018, is directly applicable in each EU Member State. The GDPR introduces a single legal framework across EU Member States, but it includes several open provisions that allow each country to restrict, specify or expand the requirements of the GDPR. This is the case with regards to the processing of genetic data, biometric data and data concerning health where Member States may maintain or introduce further limitations to the processing of these types of data.

Organisations must have a valid, legal reason to process personal data. This is called a 'legal basis' and there are six available legal basis described in Article 6. Under the GDPR, commercial companies and charitable research organisations will commonly use **'legitimate interests'** as their legal basis. However, public authorities, such as public research organisations or universities, when carrying out public tasks will use **'task in the public interest'** as their legal basis (https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/under-what-conditions-can-my-company-organisation-process-sensitive-data_en)

In order to lawfully process special category data, such as genetic data, biometric data or data concerning health, organisations must identify both a lawful basis under Article 6 and a separate condition for processing special category data under Article 9 . There are ten conditions for processing special category data in the GDPR itself, but Members States may introduce additional conditions and safeguards on the processing of genetic data, biometric data or data concerning health. Such flexibility means that any organisation processing this kind of data could be subject to different legal requirements in different countries.

Beyond these differences between Member States, there are other challenges linked more generally to the implementation of the GDPR that may have a direct impact on the processing of rare disease-related data captured in registries, biobanks, electronic health records and other databases. These include the following:

- Clarifying liability under the GDPR - who is responsible if a person figures out how to identify data that was pseudonymised in good faith?
- Operationalising the principles envisaged in the Regulation such as privacy by design and by default
- Developing standards for health data anonymization
- Clarifying the conditions to use broad consent under Recital 33 to process health data for research purposes

4.1 Patients' Perspectives on Data use and Re-Use

In recent years, research has been conducted to assess patients' perspectives on the use and reuse of their personal health-related research data. For instance, RD-Connect assembled a (disease-agnostic) panel of patient advocates, the PAC (Patient Advisory Committee). Data sharing was the topic of a recent Rare Barometer Voices survey (results to be released shortly - https://www.eurordis.org/content/eurordis-past-surveys). Such work has suggested that RD Patients are generally willing to share their health data and recognise that this is of vital importance to advance health research and healthcare, help other patients and ultimately benefit society. They have a greater incentive because data on each disease is usually very scarce and scattered, making research more challenging, and most conditions classed as rare have no cure (or even dedicated treatment). But at the same time, patients are deeply concern about privacy issues and security breaches.

Consultations and surveys suggest that RD Patients are willing to share their medical data for research as long as this is done respecting four core elements for responsible data sharing: respecting their preferences; protecting privacy and confidentiality; providing feedback on the results; and allowing patients to be part of defining the data governance and be involved in operating/managing these governance arrangements.

- **Consent is obtained respecting preferences.** Do patients have all the information they need to understand research objectives, who is going to access what data, for what purposes and under what conditions?
- **Privacy and confidentiality are protected and** mitigated through safeguards (such as ethical review, and IT solutions – privacy by design and default, security measures, data minimization, pseudonymisation…) while maintaining/respecting reasonable time frames
- **Resulting progress is communicated (feedback on the results)** Regular communication of outcomes to the patient community and the public at large should occur in a timely manner both at the aggregate and individual levels
- **Good and inclusive Health Data Governance frameworks:** In today's fast-evolving data-intensive research, while obtaining valid consent is necessary, it is not enough to restore the autonomy to individuals. Robust and transparent health data governance frameworks are required, involving patients/citizens across the data cycle and allowing them to participate actively in the collection and management of data. Clear accountability (who is responsible for misuse?) and a mechanism to redress harms should to be part of this governance framework:

5. RESULTS OF THE RARE DISEASE LITERATURE REVIEW*

**The earlier sections of this document were elaborated via research, partner expertise, and data stemming from the Resource on the State of the Art of Rare Disease activities in Europe. This final section is a summary of the results of a literature review performed by INSERM Orphanet, and is designed to highlight peer-reviewed publications which may suggest trends in this broad topic.*

Recently, **a paradigm shift may be observed when considering the place, role and attention directed towards the patient**. Indeed, it seems as if an underlying change in **The earlier sections of this document were elaborated via research, partner expertise, and data stemming from the Resource on the State of the Art of Rare Disease activities in Europe. This final section is a summary of the results of a literature review performed by INSERM Orphanet, and is designed to highlight peer-reviewed publications which may suggest trends in this broad topic.*

The emergence of a **new technological era** with the development of big data and the continuous sophistication of information and communication technologies has revolutionised many sectors, including health (Hong 2018; Belle 2015). It has both opened a field of new and **promising opportunities for the care and treatment of rare diseases, including personalised medicine,** as well as tremendous **challenges mainly linked to difficulties in finding, processing, and analysing the data and ethical issues regarding data protection.**

Firstly, a few trends can be observed when considering the process of collecting data. Our literature review identified that within the last decade, great progress has been made when looking at the number of data resources and ways of collecting data. Indeed, data for rare diseases can been found in the form of patient registries, population registries, electronic health records, as well as biobanks, each with its own characteristics and specific uses. Nevertheless, this tends to produce a situation in which these **resources multiply and divide indefinitely, creating a multitude of data silos**. Few links are made between resources and, as a result, very definite disease-specific (or disease sub-type) resources have developed, both in the public and private sector, often without a common data set (Taruscio et al. 2015; Lopes et al. 2015; Roos et al. 2017). National registries for rare diseases follow very different approaches, structures and purposes, even amongst similar and geographically proximate countries, such as European countries (Taruscio et al. 2015). This enhances the aforementioned **siloed data landscape** preventing many more general uses of the data and limiting research advances for rare diseases (Lopes et al. 2015; Roos et al. 2017).

When viewed within the context of health data, **rare disease data also tend to lack visibility in health information systems** which complicates efficient healthcare resource planning, patient management and

follow-up (Choquet et al. 2015; Marx et al. 2017). Often, codes used to define a disease vary between countries, regions and sometimes hospitals, and many rare diseases were traditionally missing from coding terminologies. This **lack of standardisation** makes it difficult to identify rare diseases and complicates the combining of data on large geographic scales, an absolute necessity in the field of rare diseases, where patients are scattered all around the world (Lopes et al. 2015; Rath et al. 2012). It also leads to 'double entries' for patients, which further complicates the task of processing the data (Choquet et al. 2015; Marx et al. 2017).

Nonetheless, when analysing the trends regarding the exploitation of the data and the informatics and bioinformatics tools designed to make sense of this huge amount of information, one can perceive **efforts across borders and across disease areas.** For instance, a **tendency towards harmonisation is appearing regarding coding practices**. Recommendations abound for routine double coding i.e. ICD-10 and Orphacodes (Marx et al. 2017) and the adoption of Common Data Elements, meaning the establishment of data elements commonly used in more than one dataset (Choquet et al. 2015; Roos et al. 2017). The overall goal of such initiatives is to break down national as well as discipline-specific barriers and easily identify patients affected by rare diseases in order to form a continuum of care across boundaries and expert centres. The general idea is to **enhance the interoperability of data and make the FAIR principles a reality**: rare diseases data should in the future be Findable, Accessible, Interoperable, Reusable (Gainotti et al. 2018; Lochmüller et al. 2018).

Another means of breaking silos observed in the literature is the use of new bioinformatics tools which allow for the combination of heterogeneous data resources and contribute to **innovative knowledge generation**. A perfect example is the l**ink made between omics and phenotypic data, creating genotype-phenotype relationships** which then enable more complete patient records and paves the way to personalised medicine (Lopes et al. 2015; Lochmüller et al. 2018). Other tools used to foster interoperability of datasets include the combinations of **semantic web, text-mining methods and ontologies** (Lopes et al. 2015).

Another significant trend in data collection is the importance and involvement of patient and family members. Patients are **solicited in their role as experts of their disease to provide data, evaluation and feedback on their experience** (Bambusch et al. 2019). This involvement prompts the emergence of **two-directional information pathways** where both patient/experiential knowledge and scientific or medical information are equally valued (Vicari and Cappai 2016). In this schema, patients become also **generators of knowledge and data,** informing research, clinical care and treatment. A direct manifestation of this trend is the development of **patient reported outcomes measures** (valuable data directly obtained from the patient about their health status or treatment without interpretation by an intermediary). These instruments help to make **patients' voices central to clinical decision-making** (Slade et al. 2018).

Finally, the collection, use and, most of all, sharing of personal and genomic data raises complex **ethical issues**. The stringent legislation of the General Data Protection Regulation implemented in May 2018, is probably the most striking example. Moreover, emphasis on the **responsibility of the data producer and user** is increasingly heightened and sanctions are currently drafted accordingly, adapting to the constant technological evolution (Takashima et al. 2018; Shabani 2016). For instance, IRDiRC partnered with the Global Alliance for Genomics and Health (GA4GH) to develop policy and guidelines around consent, data sharing and frameworks for ethical and secure data sharing, as well as promoting standards for nomenclature (Lochmüller et al. 2017).

Our literature review suggests a general promotion for the design and implementation of policies related to data protection, security and privacy with the **need to find a balance between data sharing and data protection** (Takashima et al. 2018) so as not to hinder scientific advances. The focus and importance on the **anonymisation of data** is a sign of such consideration (Oprisanu and De Cristofaro 2018). Furthermore, privacy seems to have become a central concern and **more attention is paid to patients' opinions and their perspectives on data and biomaterial sharing** (McCormack et al. 2016).

Finally, a quite novel trend which, among other things, could bring a solution to privacy concerns regarding data sharing, is the use of **blockchain technology**. This can be defined as an ever-growing list of records linked using cryptography and containing information that can be simultaneously used and shared within a large decentralized, publicly accessible network. Indeed, t**his system could ensure patients' ability to retain ownership on their data**, one of the core elements for the respect of privacy according to some experts (Angeletti et al. 2017; Terry and Terry 2011) and hence provides an **innovative way to improve the intelligence of healthcare systems while keeping patient data private** (Yue 2016).

REFERENCES FROM THE RARE DISEASE LITERATURE REVIEW

FULL LIST OF ARTICLES / PUBLICATIONS FOUND IN THE LITERATURE REVIEW:

- https://docs.google.com/spreadsheets/d/1SRXASsFiD9sdQz286SVo860XdTpGaOIncyjIhGphULI/edit?usp=sharing

SELECTED BIBLIOGRAPHY OR KEY PUBLICATIONS

- Angeletti, F., Chatzigiannakis, I., Vitaletti, A., 2017. The role of blockchain and IoT in recruiting participants for digital clinical trials, in: 2017 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM). Presented at the 2017 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), IEEE, Split, pp. 1–5. https://doi.org/10.23919/SOFTCOM.2017.8115590
- Baumbusch, J., Mayer, S., Sloan-Yip, I., 2019. Alone in a Crowd? Parents of Children with Rare Diseases' Experiences of Navigating the Healthcare System: J Genet Couns 28, 80–90. https://doi.org/10.1007/s10897-018-0294-9
- Belle, A., Thiagarajan, R., Soroushmehr, S.M.R., Navidi, F., Beard, D.A., Najarian, K., 2015. Big Data Analytics in Healthcare. BioMed Research International 2015, 1–16. https://doi.org/10.1155/2015/370194Choquet, R., Maaroufi, M., de Carrara, A., Messiaen, C., Luigi, E., Landais, P., 2015. A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research. Journal of the American Medical Informatics Association 22, 76–85. https://doi.org/10.1136/amiajnl-2014-002794
- Gainotti, S., Torreri, P., Wang, C.M., Reihs, R., Mueller, H., Heslop, E., Roos, M., Badowska, D.M., de Paulis, F., Kodra, Y., Carta, C., Martìn, E.L., Miller, V.R., Filocamo, M., Mora, M., Thompson, M., Rubinstein, Y., Posada de la Paz, M., Monaco, L., Lochmüller, H., Taruscio, D., 2018. The RD-Connect Registry & Biobank Finder: a tool for sharing aggregated data and metadata among rare disease researchers. Eur J Hum Genet 26, 631–643. https://doi.org/10.1038/s41431-017-0085-z
- Lopes, P., Silva, L.B., Oliveira, J.L., 2015. Challenges and Opportunities for Exploring Patient-Level Data. BioMed Research International 2015, 1–11. https://doi.org/10.1155/2015/150435
- Marx, M.M., Dulas, F.M., Schumacher, K.M., 2017. Verbesserung der Sichtbarkeit seltener Erkrankungen in Gesundheitssystemen durch spezifische Routinekodierung. Bundesgesundheitsbl 60, 532–536. https://doi.org/10.1007/s00103-017-2534-9
- McCormack, P., Kole, A., Gainotti, S., Mascalzoni, D., Molster, C., Lochmüller, H., Woods, S., 2016. 'You should at least ask'. The expectations, hopes and fears of rare disease patients on large-scale data and biomaterial sharing for genomics research. Eur J Hum Genet 24, 1403–1408. https://doi.org/10.1038/ejhg.2016.30

- Musterman, M., Placeholder, P., n.d. Liang Hong, Mengqi Luo, Ruixue Wang, Peixin Lu, Wei Lu*, Long Lu. Data acquisition 23.
- Oprisanu, B., De Cristofaro, E., 2018. AnoniMME: bringing anonymity to the Matchmaker Exchange platform for rare disease gene discovery. Bioinformatics 34, i160–i168. https://doi.org/10.1093/bioinformatics/bty269
- Rath, A., Olry, A., Dhombres, F., Brandt, M.M., Urbero, B., Ayme, S., 2012. Representation of rare diseases in health information systems: The orphanet approach to serve a wide range of end users. Hum. Mutat. 33, 803–808. https://doi.org/10.1002/humu.22078
- RD-Connect consortium, NeurOmics consortium, EURenOmics consortium, Lochmüller, H., Badowska, D.M., Thompson, R., Knoers, N.V., Aartsma-Rus, A., Gut, I., Wood, L., Harmuth, T., Durudas, A., Graessner, H., Schaefer, F., Riess, O., 2018. RD-Connect, NeurOmics and EURenOmics: collaborative European initiative for rare diseases. Eur J Hum Genet 26, 778–785. https://doi.org/10.1038/s41431-018-0115-5
- Roos, M., López Martin, E., Wilkinson, M.D., 2017. Preparing Data at the Source to Foster Interoperability across Rare Disease Resources, in: Posada de la Paz, M., Taruscio, D., Groft, S.C. (Eds.), Rare Diseases Epidemiology: Update and Overview. Springer International Publishing, Cham, pp. 165–179. https://doi.org/10.1007/978-3-319-67144-4_9
- Slade, A., Isa, F., Kyte, D., Pankhurst, T., Kerecuk, L., Ferguson, J., Lipkin, G., Calvert, M., 2018. Patient reported outcome measures in rare diseases: a narrative review. Orphanet J Rare Dis 13, 61. https://doi.org/10.1186/s13023-018-0810-x
- Takashima, K., Maru, Y., Mori, S., Mano, H., Noda, T., Muto, K., 2018. Ethical concerns on sharing genomic data including patients' family members. BMC Med Ethics 19, 61. https://doi.org/10.1186/s12910-018-0310-5
- Taruscio, D., Vittozzi, L., Choquet, R., Heimdal, K., Iskrov, G., Kodra, Y., Landais, P., Posada, M., Stefanov, R., Steinmueller, C., Swinnen, E., Van Oyen, H., 2015. National Registries of Rare Diseases in Europe: An Overview of the Current Situation and Experiences. Public Health Genomics 18, 20–25. https://doi.org/10.1159/000365897
- Terry, S.F., Terry, P.F., 2011. Power to the People: Participant Ownership of Clinical Trial Data. Science Translational Medicine 3, 69cm3-69cm3. https://doi.org/10.1126/scitranslmed.3001857
- Vicari, S., Cappai, F., 2016. Health activism and the logic of connective action. A case study of rare disease patient organisations. Information, Communication & Society 19, 1653–1671. https://doi.org/10.1080/1369118X.2016.1154587
- Xue, X., Wang, H., Jin, D., Li, M., Jiang, W., 2016. Healthcare Data Gateways: Found Healthcare Intelligence on Blockchain with Novel Privacy Risk Control. J Med Syst 40, 218. https://doi.org/10.1007/s10916-016-0574-6

Rare2030

Foresight in Rare Disease Policy

The health of 30 million people living with a rare disease in Europe should not be left to luck or chance. The Rare 2030 foresight study prepares a better future for people living with a rare disease in Europe by gathering the input of a large group of patients, practitioners and key opinion leaders to propose policy recommendations.

Since the adoption of the Council Recommendation on European Action in the field of Rare Diseases in 2009, the European Union has fostered tremendous progress to improve the lives of people living with rare diseases. Rare2030 will guide a reflection on rare disease policy in Europe through the next ten years and beyond.

PARTNERS

The Rare2030 project is co-funded by the European Union Pilot Projects and Preparatory Actions Programme (2014-2020). This leaflet is part of the pilot project PP-1-2-2018-Rare 2030. The content represents the views of the author only and is his/her sole responsibility; it cannot be considered to reflect the views of the European Commission or any other body of the European Union.

This document was elaborated by Rare 2030 WP4, under the leadership of Newcastle University and with input from all partners. The document should be referenced as follows: Hedley, V., Kole, A., Rodwell, C., and Simon, F. (2019) *Rare 2030 Knowledge Base Summary on Data Collection and Utilisation for Rare Diseases* https://www.rare2030.eu/our-work/