Project 101232916 - EUCMTSC

Deliverable D.3.1.

Outline for a data-sharing framework in conformity with GDPR and EHDS, and easily accessible for researchers

Maike Dohrn[1], Alexandre Hoyau[2], Ingolf Pernice[3], and Wolfgang Pernice[4]

[1] RWTH Aachen University Hospital, Germany; ECRA

[2] CMT France, ECRA

[3] ECMTF

[4] Columbia University Medical Center, New York, USA; ECMTF

Key words: Antwerp Conference, ECMTSC, Data sharing, health data, rare diseases, neuromuscular diseases, CMT, GDPR, EHDS, patient registries, ECMTF, ECRA, JARDIN, ERN Euro-NMD, GENESIS, BNDMR, Gene Matcher, GnomAD, patient ownership, trust, ENMC

ECRA's perspective on data sharing: where are we now?

Data sharing is on everyone's lips – yet, the biomedical data landscape remains highly fragmented. Broad sharing and consolidation of both medical health data and research-level datasets is critical to biomedical research progress, and especially so for rare diseases like CMT. Legitimate economic, competitive, legal, and security implications to data sharing have led to a diversity of siloed patient registers, health- and biomedical research databases of various shapes and sizes that generally limit access to select groups of investigators. Absent solutions to overcome these barriers, major opportunities for scientific advancement, as would be enabled by broad-scale synergistic combination of genetic, functional omics, and clinical data, in combination with new technologies including Artificial Intelligence (AI), may remain out of reach for CMT. On the initiative of patient organizations (ECMTF) and within the frame of the Antwerp Conference (ECMTSC), ECRA has taken on the task of leveraging assembly of stakeholders across all domains of the CMT biomedical care and research ecosystem, including basic scientists, clinicians, and patient representatives, to shed light on the opportunities and challenges of data sharing from all angles. It is aware of the new EHDS-Regulation the implementation of which may be anticipated within the applicable limits.

Motivation: *Why do we want to share data, specifically, in the CMT field?*

Worldwide, CMT is considered a rare disease. Still, we estimate that 300,000 affected individuals live in Europe as of now, with a potentially significant number of unreported cases on top. With more than 100 genes causing CMT if mutated, CMT exhibits an extremely long-tailed distribution of subtypes many of which meet the definition of ultra-rare conditions, affecting <1:50,000 people.

ECRA members strongly believe that genetic and phenotypic diversity can only be sufficiently addressed through an open, collaborative, and multidisciplinary network. Specific advantages

of data sharing include more efficient genotype-phenotype matching across centers, which subsequently allows for (1) the establishment of evidence-based gene-disease relationships, (2) novel gene discoveries, (3) expedited diagnosis, (4) accelerated mechanistic and translational research, and (5) enhanced recruitment for clinical trials.

Data Types: *What exactly do we want to share in the CMT field?*

To enable optimized collaboration and effectiveness, ECRA suggests creating an infrastructure to facilitate free access and sharing of data on patient genotypes, functional omics data including e.g. RNA sequencing or high-content imaging data, clinical phenotypes, and the availability of tissue samples for research purposes.

Framework scope: *What scope will data sharing be operated on?*

ECRA members are committed to pursuing solutions that can support international data sharing on a global rather than European scale only. However, given that EU data protection regulations have at least been perceived to impose the most restrictive legal context for international data sharing, developing pertinent GDPR-compliant pathways and frameworks to support international sharing of covered health and biomedical information presents a primary objective for ECRA.

Framework structure: *What structure should the data sharing framework have?*

For the abovementioned reasons (rareness and diversity of CMT), a network structure may be most versatile and promising, if it is open and designed to foster international collaboration for maximum outcome. However, should inherent challenges of network structures prove difficult to reconcile with these objectives, a new approach in shape of a patient-owned, independent, and public health-data base may be a possible alternative to be considered.

Challenges: *Why is it so difficult to align centers to leap into data sharing?*

ECRA is aware of the following key challenges:

1) Any data sharing strategy must be compliant with both EU level data protection regulations (GDPR), and relevant national law, and needs authorization by pertinent local Ethics and/or regulatory committees (e.g. Institutional Review Boards) for each participating investigator's institution(s). Upcoming EHDS-Regulation may go a long way towards harmonizing relevant standards and procedures that could reduce real and perceived regulatory barriers to data sharing.

2) Across domains, data exist in heterogeneous formats. Adequate technical metadata may not always be readily available, limiting the re-usability and interoperability of datasets. Significant curation, annotation, or re-processing may be required to enable effective data sharing.

3) The costs of hosting a data sharing platform can be significant, especially in case of success. In search for a sustainable long-term funding model, a fee-for-access model is intuitive, yet threatens the goals of a data sharing solution as we envision it by imposing unequal financial barriers for academic researchers dependent on their funding.

4) Sharing data comes with a competitive advantage for the receiver, which means that a fair and equitable system of access and reward is needed to prevent unfair exploitation and to incentivize sharing of data by data generators.

5) Hosts and/or administrators of data sharing platforms inevitably enjoy root-access privileges (i.e. access to *all* data) that can strongly disincentivize other data generators from contributing their own data to the platform, especially if the hosts are also engaged in relevant research.

To overcome these challenges for CMT research in Europe and beyond, ECRA has defined the following **worksteps**:

1. To involve basic scientists, clinicians, and patients to increase transparency and build trust
2. To assess the existing infrastructure and analyze the unmet needs
3. To develop strategies on how to best share data for the benefit of science – and patients

For **workstep 1**, the Antwerp meeting gave us the opportunity to hear presentations on existing data sharing frameworks and/or platforms (e.g. ERN Euro-NMD) and stimulated discussions that form the basis of this report. In this context, we performed a SWOT analysis (strengths, weaknesses, opportunities, threats), as displayed in the attached overview chart. To further inform ECRA's perspective on data sharing, we decided to collaborate closely with the EU-funded JARDIN initiative (Joint Action on Integration of ERNs into National Healthcare Systems – contact persons: Dr. Anne-Sophie Lapointe, Alexandre Hoyau) and to extrapolate suggestions on best practices for data sharing for the CMT community.

In **workstep 2,** we will assess more deeply the unmet needs across existing infrastructure. Among institutions/ data sharing frameworks, we will focus our review on:

1. The GENESIS network (University of Miami, US): GENESIS is a centralized genomic data interpretation and sharing platform that hosts the largest accessible collection of genomes in CMT. It enables users to upload raw genomic sequencing data (panel, exome, genome, long-read), to be preprocessed via standardized pipelines, to interpret variants in a user-friendly fashion (GUI-based; no bioinformatics background required), and to query against a growing set of "publicly available" genomes of patients with CMT and related conditions. GENESIS is linked with the AI-tool MAVERICK that assists with the interpretation and prioritization of genetic variants and has already contributed to a significant number of new gene discoveries. Several ECRA members are already part of the GENESIS network.
2. Gene Matcher (Broad Institute, US): Gene Matcher is a globally used distributed rare variant discovery tool that enables investigators to identify patients with matching genetic variants, without the need to share phenotypic (or other) information a priori. Upon a match, investigators are notified and invited to connect to initialize scientific collaborations as pertinent.
3. GnomAD (Broad institute, US): GnomAD is a centralized and extensively used genetic data sharing resource that hosts *freely and publicly accessible* aggregate-level exome and genome reference data (allele frequencies), a pillar of human genomic research over the last decade.
4. Euro-NMD: For curated phenotype data and gene-disease associations, the European Reference Network (ERN) has established a registry (Euro-NMD) with a specific focus on hereditary neuropathies. In a shared minimal data set, investigators aggregate and correlate genotypes (pre-interpreted, no raw data) and phenotypes and link these records to the availability of tissues or other patient materials. Several ECRA members are already part of Euro-NMD.
5. BNDMR (French National Database): BNDMR is another example for a rare disease patient registry; it contains data sets from 1.4 million rare diseases patients, including ~8000 individuals with CMT.
6. INC: The Inherited Neuropathy Consortium (INC) is a primarily US-based clinical research consortium that has driven and coordinated collaborative research, investigator training, and the development of key resources and tools for the field of CMT over the last

10-15 years. Being the leading CMT research consortium, the INC has collected major compendiums of natural history data, genetic data (GENESIS), biomarker data, as well as mechanistic data across large cohorts of CMT patients and supported recruitment for clinical trials (NCT01733407, NCT05397665). The INC presents an archetype Rare Disease Clinical Research Consortium (RDCRN) as originally sponsored by the U.S. National Center for Advancing Translational Sciences (NCATS).

The objective of our review is to identify key benefits and limitations relevant to the CMT field for each of these examples as to inform the policy, technical, and economic principles for a general data sharing solution in Europe.

In **workstep 3,** ECRA centers with successful IRB and/or GDPR approval for data sharing (e.g. with GENESIS or Euro-NMD) will share their forms and vota as a template for other centers. This will decrease barriers and save time to initiate participation. Workstep 3 has already been initiated in Antwerp, where ECRA members have begun to exchange files and experiences. This rather practical step promises to reveal so far unknown barriers and challenges of data sharing that may be overcome with innovative instruments and procedures to be developed.

The patients' role: *What can "patients as partners" in CMT-research effectively contribute?*

ECRA's data sharing **strategy** is solidly based on the concept of **patient ownership**. Across most jurisdictions worldwide (E.g. EU law, GDPR, articles 4, 13, 14, 15, 16, 18, 20, 21), patients have fundamental rights over their data. From a legal point of view, researchers are allowed to use these data under certain conditions, including informed consent. Consent means giving permission to using specific data, which is opposed to transferring rights or property. The Antwerp meeting enabled ECRA members to discuss how patients, clinicians, and basic scientists perceive their specific roles, and how these are being anchored in EU law. For CMT data sharing, patient ownership means that patients are aware of their right to access (GDPR article 15) and move (GDPR article 20) their data. Therefore, they are entitled to collect their data, such as molecular genetic sequencing results, and transfer them to a center that is qualified to share the above upon consent in a research network – or a public body – as it might prove to be more appropriate.

Building trust means creating an atmosphere in which patients feel safe to entrust researchers with their data and, at the same time, researchers themselves benefit from sharing these data with an expert community rather than keeping them at their own centers for good. Strong incentives can be shared publications and grants, and faster progress in one's own research. For this, ECRA brings together: (1) clinicians and scientists who have experience with both successes and challenges of data sharing, (2) those who are collecting data and are committed to further advancing fair and reciprocal data sharing solutions, and (3) patients who are motivated to promote efficient CMT research by sharing their data with qualified researchers.

Conclusion

Based on the present assessment, ECRA in close cooperation with ECMTF will continue to discuss in-depth how to solidify concrete measures to further 1) address challenges, 2) increase trust, 3) and empower patients, within a system that serves as a magnet for outside organizations to join in. We strongly believe that for this concept, the Antwerp Conference created an indispensable nucleus.