

Meeting Report

How Patient Organizations Can Drive FAIR Data Efforts to Facilitate Research and Health Care: A Report of the Virtual Second International Meeting on Duchenne Data Sharing, March 3, 2021

Nawel van Lin^{a,e,*}, Georgios Paliouras^{b,c}, Elizabeth Vroom^{a,d}, Peter A.C. 't Hoen^c and Marco Roos^f

^a*Duchenne Parent Project, Veenendaal, The Netherlands*

^b*Duchenne Data Foundation, Veenendaal, The Netherlands*

^c*National Centre for Scientific Research, “Demokritos”, Agia Paraskevi, Greece*

^d*World Duchenne Organization UPPMD, Veenendaal, The Netherlands*

^e*Center for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands*

^f*Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands*

Pre-press 29 July 2021

Abstract.

Background: For patients with rare diseases such as Duchenne and Becker muscular dystrophy (DMD/BMD), access to their health data is key to being able to advocate for themselves and be in control of their care. Since 2018, the DMD/BMD patient community has been committed to making DMD/BMD-related data FAIR, i.e., Findable, Accessible, Interoperable, and Reusable. On March 3, 2021, the second international meeting on FAIR data sharing for DMD/BMD was held virtually.

Objective: The aim of this meeting report is to summarize the presentations and discussions of the meeting.

Methods: During this meeting, the progress of FAIRification efforts since the first international meeting in 2019, new developments, stakeholder perspectives, and experiences from implementing FAIR data principles in practice were presented and discussed.

Results: Over 120 attendees representing various stakeholder groups (ie, patient organizations, clinicians, clinical and academic researchers, pharmaceutical companies, regulators, and EU organizations) from 22 countries participated in the meeting. This meeting report summarizes the presentations and discussions from the meeting, provides an overview of the key lessons learned since the first meeting, and outlines the next steps.

Conclusions: Patient organizations are key drivers of the FAIRification process in practice and dialogue with stakeholders is critical to success.

Keywords: Duchenne muscular dystrophy, Becker muscular dystrophy, rare diseases, FAIR data principles, data sharing, data visiting, big data, data analytics

*Correspondence to: Nawel van Lin, Landjuweel 34, 3905PG Veenendaal, The Netherlands. Tel.: +31 20 627 5062; E-mail: nawel.vanlin@worldduchenne.org.

INTRODUCTION

They say knowledge is power. For patients with rare diseases such as Duchenne and Becker muscular dystrophy (DMD/BMD) and their caregivers, knowledge means accurate and timely diagnosis, effective health care and treatment, and an independent and productive life (“actionable knowledge”). Knowledge also means successful advocacy and empowerment.

The reality for patients with DMD/BMD is that from the moment of diagnosis, they are seen by many different medical specialists and healthcare professionals. Consequently, a wide range of health data is produced, encompassing results from diagnostic tests, genetic tests, and physical examination; data related to activity of daily living (ADL) such as the use of assistive devices and participation; patient-reported outcomes (PROs) in response to received care and treatment; treatment history; and participation in clinical trials. These data are often kept in many different data systems and their (re)use is prevented by prevailing silo mentalities and issues such as ownership, control, legislation, data protection, and data security. Data are formatted for one type of use and understood by a single user or a small group of users, which impedes the data’s use beyond this scope. This limitation interferes with the discovery of new diagnostics, treatments, and health care policies to benefit patients with DMD/BMD. It also hampers the use of data by patients themselves.

The adoption of FAIR data practices to make data Findable, Accessible, Interoperable and Reusable for humans and computers by all stakeholders who build, manage, and own these data systems can address this usage problem. Making data readable by both humans and machines (ie, machines knowing what our data mean) can be achieved by special software and “knowledge representation” technology. This enables efficient and real-time analysis across multiple data sources without moving data [1]. It also allows us to specify and control (1) what type of data queries can be made within a given source of data and by whom, and (2) what data (e.g., the results of queries) can be seen outside of the source. Specifically, an algorithm may have permission to calculate summary data from all pseudonymized data inside a data source, but only deidentified and aggregated summary data are allowed to leave the source for human inspection. Applying knowledge representation technology during a FAIRification process makes conditions for access transparent and

assessable and facilitates compliance with the General Data Protection Regulation (GDPR; for more on this see Landi et al. [2] and Brewster et al. [3]). Consequently, FAIR data can be used appropriately by researchers, health care providers, regulators, family members, and patients to accelerate discoveries for early diagnosis and innovative treatments. Furthermore, FAIR data enables treatments to be personalized to the needs of individual patients. This is ultimately what patients with DMD/BMD want [4].

In 2018, members of the DMD/BMD patient community (the Dutch Duchenne Parent Project, World Duchenne Organization, and the Duchenne Data Foundation) came together to address accessibility issues with respect to their own health data. Patients raised concerns about the barriers that prevent researchers, health care providers, policy makers, and regulatory authorities from (re)using DMD/BMD related data for the benefit of all those suffering from DMD/BMD. This resulted in two actions.

First, the DMD/BMD patient community defined their position regarding DMD/BMD-related data. Namely, they stated their support for (1) optimal (re)use of data collected from different DMD/BMD data sources; (2) collection of PROs; (3) access to placebo data from clinical trials; (4) returning clinical trial data to participants; (5) giving patients the right to decide how their data are used; and (6) viewing GDPR as a catalyst that allows patients who are willing and interested to share their data [4, 5].

Second, the DMD/BMD patient community made a commitment to make DMD/BMD-related data FAIR. This commitment led to the first international meeting on FAIR data sharing for DMD/BMD in 2019, which aimed to inform the stakeholders in the community about FAIR data principles and the benefits of implementing them in practice [4, 6]. Following this meeting, DMD/BMD patient organizations established collaborative relationships with FAIR data experts, and became involved in FAIR-related projects (e.g., Brain Involvement in Dystrophinopathies [BIND], the Duchenne Data Platform, and the European Reference Network for neuromuscular diseases [EURO-NMD] registry hub).

On March 3, 2021, the second international meeting on DMD/BMD data sharing took place virtually. This meeting reflected the DMD/BMD patient community’s continuing belief in the power of linking health data. The purpose of the meeting was to provide an update on the progress of

FAIR data implementation projects since 2019, to inform the community of the latest FAIR developments, share stakeholder perspectives on and experiences in implementing FAIR in practice, and discuss future steps. Speakers and attendees included representatives of DMD/BMD patient organizations, the European Medical Agency (EMA), Rare Diseases Europe (EURORDIS), and the European Joint Research Programme on Rare Diseases (EJP-RD); clinicians and clinician-researchers; pharmaceutical companies; and FAIR data experts, researchers, and consultants. There were 128 attendees from 22 countries (i.e. Argentina, Belgium, Brazil, Canada, Czech Republic, Chili, France, Germany, Greece, Netherlands, USA, UK, Israel, Italy, Serbia, Spain, Turkey, Kenya, Mexico, Poland, Ukraine, and Norway). A video recording of the meeting is available online [7].

The aim of this report is to summarize the speakers' presentations and discussions among speakers and attendees during the meeting. By doing so, we hope to engage more stakeholders in making DMD/BMD data FAIR, accelerate discoveries to benefit the health and wellness of patients with DMD/BMD, and facilitate efforts by patient organizations for other rare diseases. To facilitate readability, we divided the report into the following six thematic sections that were discussed across the sessions of the original meeting schedule: (1) new initiatives in DMD/BMD data collection; (2) implementing FAIR data principles in practice; (3) building a solid ecosystem for DMD/BMD data sharing; (4) lessons learned; (5) next steps; and (6) final remarks. Also, as FAIR data sharing is a relatively new development, where concepts may be unfamiliar or too technical, we added some clarification. There were no controversial discussions where opinions were conflicting. For an overview of the speakers and their presentations in the chronological order during the meeting, please refer to Supplementary Table 1.

NEW INITIATIVES IN DMD/BMD DATA COLLECTION

Since 2019, three initiatives to develop DMD/BMD-related data collection systems based on FAIR data principles have been launched. Two initiatives are related to patient-empowered registries by DMD/BMD patient organizations (i.e., the Parent Project Muscular Dystrophy in the USA and the Duchenne Parent Project in The Netherlands). The third is a data repository to connect multiple sources

that store DMD/BMD-related data created by the Duchenne Data Foundation. The speakers were Ryan Fisher (Parent Project Muscular Dystrophy, USA), Elizabeth Vroom (World Duchenne Organization, The Netherlands), and Georgios Paliouras (Duchenne Data Foundation, Greece).

Patient-empowered registries

Patient-empowered registries allow patients and their caregivers to enter and access their health data as they wish. In addition, patients may determine which data may be accessed and used by others. Examples of data that may be captured in patient-empowered registries include data from genetic studies, medical visits to specialists and paramedical professionals, eHealth medical record data, clinical trial data, wearables, PROs, and ADL.

Ryan Fischer described how the efforts of the Parent Project Muscular Dystrophy in the USA to make DMD/BMD data FAIR have been facilitated by new federal legislation (The 21st Century Cures Act) that defined interoperability of data and prohibited information blocking [8–10]. Currently, two consortiums (Duchenne Regulatory Science Consortium and the Collaborative Trajectory Analysis Project) aim to take natural history data and placebo-arm data from clinical trials out of their silos and develop DMD/BMD progression models. Furthermore, the Parent Project Muscular Dystrophy (in collaboration with Prometheus Research) is developing a data hub (Duchenne Outcomes Research Interchange) where data from its patient registry (e.g., PROs), eHealth records, and clinicians (i.e., postmarket data on therapeutics) can be combined. This will enable clinicians to create deidentified data sets to compare data across DMD/BMD clinics and gain insight into progression in DMD/BMD [11].

Elizabeth Vroom presented how since the 2019 meeting, the Duchenne Parent Project in The Netherlands established a dedicated technical team and started the process of developing its own patient-empowered registry (Duchenne Data Platform) based on the premise that patients own their own data, should collect data that are relevant to them (such that the platform can complement priorities defined by clinicians), and should support data to be as usable as possible under defined conditions [12]. In 2020, additional developments took place to transform the Duchenne Data Platform into a FAIR-enabling data source guided by FAIR data principles and FAIR experts. Within the data platform, patients

have their own so-called virtual data locker that they can access through their cell phone or computer. By means of dynamic informed consent, patients can determine who has access to their data and when. Data elements inside the platform were made interoperable and machine-readable by annotating them with terms from global ontologies, partly reusing a knowledge representation model defined in a FAIR-related project by the EJP-RD. Consequently, third parties such as regulators, companies and research institutes are able to “visit” the data contained in the registry according to FAIR protocols. Moreover, the data platform offers patients the opportunity to make their own queries as “citizen researchers”. Once the FAIR transformation of the platform is completed, it can be linked to other FAIR data sources such as the Duchenne Data Repository. Once the platform and the Duchenne Data Repository are linked, data-intensive research across multiple DMD/BMD-relevant resources will be possible and more efficient. Furthermore, in order to include patients with DMD/BMD from around the world and reduce the need for duplicate efforts by other DMD/BMD patient organizations, the soon-to-be FAIR Duchenne Data Platform has been translated into English, Spanish, and Dutch. The platform has also been designed with flexibility that allows the inclusion of country-specific items.

The Duchenne data repository

Georgios Paliouras described the creation of a FAIR by design Duchenne Data Repository, which began in 2020. The aim is to foster the convergence and integration of DMD/BMD-related data and metadata (information about the data) from all existing digital data sources around the world. Through virtual extensions and implemented interoperability to link with other DMD/BMD databases, users will be able to create new “virtual” data sets with data from multiple sources without physically moving data from the original sources. In addition, the user interface of the repository is designed as a collaborative environment, which will enable multiple users to work together, communicate, and share newly created data sets.

Summary of DMD/BMD patient community efforts

In summary, the DMD/BMD patient community has turned their commitment to FAIR data principles

into action as demonstrated by the development of FAIR patient-empowered data platforms and a FAIR data repository. By (re)designing patient-empowered data platforms according to FAIR data principles, data from these platforms may be integrated and analyzed to (1) improve the DMD/BMD patient community’s understanding of the different standards of care being offered to patients with DMD/BMD and inform health care policies; (2) provide insight into the natural history of DMD/BMD (including patient-relevant aspects such as the use of assistive devices and participation) and the inclusion of PROs as end points in clinical studies for new therapeutics; (3) recruit patients for clinical studies; (4) understand the real-world impact of therapeutics; and (5) provide DMD/BMD advocates with relevant data for discussions with regulators and policy makers.

IMPLEMENTING FAIR DATA PRINCIPLES IN PRACTICE

During meeting, the following leading FAIR data experts shared insights: Barend Mons (CODATA, The Netherlands), Peter-Bram’t Hoen (Radboudumc, The Netherlands), and Marco Roos (LUMC, The Netherlands). Also, FAIR data consultants Bruna Dos Santos Vieira (Radboudumc, The Netherlands), Mark Wilkinson (BBVA-UPM, Spain), Mike Rose (CABI, UK), and Tony Burdett (FAIRplus, UK), and a FAIR project manager, Nawel van Lin (Duchenne Parent Project, The Netherlands) shared insights from experiences of implementing FAIR-related projects in practice (Supplementary Table 1). These insights related to required culture changes and recent/ongoing FAIRification projects.

Culture change to facilitate successful implementation

Marco Roos proposed that while endorsement of the FAIR principles by global organizations such as International Rare Diseases Research Consortium (IRDiRC) is an important step in facilitating implementation in practice, successful implementation ultimately depends on transforming the current silo-based data management culture. Such transformation needs to be triggered by so-called game changers who are committed to disrupting the status quo and championing a new endgame. In this new endgame, stakeholders’ data usage is characterized by greater efficiency and creativity and addresses

patient priorities. Given that stakeholders who are served well by the status quo are often resistant to change, the Rare Diseases Global Open FAIR implementation network (RDs GO FAIR) considers patient organizations to be prime candidates to act as game changers. In practice, our experience in the DMD/BMD community is that “FAIR project managers” employed by patient organizations are critical game changers. FAIR project managers serve as a critical link between (1) managers of data that should be reused (i.e., patient organizations and database software providers) and (2) data scientists who can make machines understand data for reuse. Moreover, through the FAIR project manager, patient organizations and thus patients, not only have a seat at the table, but drive the changes to ensure that their data can be reused as they wish. FAIR project managers also have a valuable role in increasing awareness about FAIR among stakeholders. For example, they can advocate to decision makers that the decision on which software provider to hire should be based on the condition that the software provider follows FAIR data principles.

Mike Rose identified another required culture change from experiences involving donor-funded data management projects such as those carried out by the intergovernmental, not-for-profit Centre of Agriculture and Bioscience International (CABI) [13]. When a funding agency decides to invest in data-based projects, the approach often involves focussing on the technology first, then looking at the workflow, and lastly, considering users of the system. In the end, these projects fail because this approach results in a system where data are left unused as the intended users cannot access data due to context-specific issues. The solution lies in reversing the process and adopting a “people first” approach to understand the unique background and context of the targeted users of the data management system; the workflow specifications and design of the technology follow thereafter. This means beginning by asking questions such as who will use the data, where is the data coming from, what are the contexts and beliefs of the people who have to share data, how do people interact with each other, how do they understand FAIR data principles, and what do the principles mean to them as individuals. The answers will shape the rules and processes that get set and help ensure that the implemented data management system reflects what the people want, which increases the chances of a project succeeding and leading to a return-on-investment.

Insights from FAIRification projects

First, Barend Mons presented insights from creating the Virus Outbreak Data Implementation Network (VODAN) at the onset of the COVID-19 pandemic [14]. This initiative has led to widespread recognition of the value of a FAIR-based data management network as VODAN has enabled researchers around the world to access data that are stored in multiple locations at any time. The FAIR-based data management network can be conceptualized as a federation of virtually linked data management systems, each of which acts as a “data station”. Communication between the data stations occurs by means of computer-readable messages that metaphorically act as “trains” [15]. When a user has a question, a train with a computer-readable query visits each data station and assesses its metadata (information about the station and its data) to find, access, reuse relevant data, and compile an answer from one or more stations. Each data station is built with specifications as to what types of query-related algorithms are permitted and which results can be returned. As such, data remain at the source and are shared by visiting. Furthermore, the VODAN project, as well as others, are conducting proof-of-concept studies regarding dynamic informed consent by the patient. Findings from these studies will provide new insights about dynamic informed consent for patient-empowered registries.

Second, Peter-Bram ‘t Hoen shared insights from establishing a FAIR-based network of data for health care providers and expert centers on neuromuscular diseases at the European level through European Reference Networks (ERNs). ERNs are virtual networks of reference centers that patients can access through their local physicians to obtain specialist support for the diagnosis and treatment of complex or rare diseases and conditions (i.e., EURO-NMD). To establish this network, EURO-NMD is creating a data hub that links existing patient-related registries as a federation in order to answer questions such as “What are the differences in age at loss of ambulation as a result of steroid use in DMD/BMD patients across EU countries?” Connecting existing registries as a federation helps mitigate the redundancy problem associated with building a central database and moving all data into it. The aforementioned Duchenne Data Platform is one of the registries that will be linked to the EURO-NMD registry through FAIR protocols. The DMD/BMD patient community efforts can also benefit from experiences with federated

machine-learning across FAIR data sources, such as in the Personal Health Train initiative; [16] the adoption of FAIR data principles by international infrastructures, such as ELIXIR, the European Open Science Cloud, and the European Health Data Space; [17, 18] the availability of reusable (meta)data standards and policies at www.FAIRsharing.org, [19] including a semantic model for common data elements in rare disease registries; and the development of machine-readable informed consent, generalized authentication and authorization mechanisms.

Third, Bruna Dos Santos Vieira discussed insights from a *de novo* FAIRification project. The European reference network for rare multisystemic vascular diseases created and launched a data registry called VASCA that was built using a *de novo* FAIRification process. With a *de novo* approach, data are made FAIR automatically upon collection within an electronic data capture system. Key aspects to successfully creating VASCA were the careful selection and definition of data variables, and a multidisciplinary team that included domain experts who could help semantic data model specialists interpret the meaning of the chosen data variables. Through VASCA, nine expert centers for rare multisystemic vascular diseases across Europe are connected as a federation; all nine have their own database where they enter, store, and maintain control of their data. The experience of creating a FAIR-based data registry and links to relevant resources such as a codebook for common data elements (CDEs) and the EJP-RD semantic model have been published and available to reuse by DMD/BMD patient organizations and other DMD/BMD stakeholders [20–24].

Fourth, Tony Burnett shared insights from FAIRplus, a European-level initiative by 22 partners from academia, companies and the European Federation of Pharmaceutical Industries and Associations (EFPIA). FAIRplus aims to implement FAIR data principles in the pharmaceutical industry and exploit the benefits of a federated approach to data management, sharing, and analysis. Specifically, FAIRplus aims to (1) create a practical guide on FAIRification best practices; (2) increase the FAIR levels of 20 Innovative Medicines Initiative (IMI)-sponsored projects and internal EFPIA data sets; (3) identify FAIRification approaches that work in the real world, including how to assess the FAIRness level of data at the start of a project; and (4) change the current data management culture in the pharmaceutical industry by ensuring that data collected in projects conducted by EFPIA and the IMI are “born” FAIR. FAIRplus

also follows a people-first, agile, and reiterative approach that involves identifying the FAIRification goal, evaluating the starting point, designing change to close the gap between starting point and FAIRification goal, implementation of the change, and re-evaluation. Experiences from the FAIRplus initiative have uncovered the following needs: good case studies that demonstrate the return on investment of FAIRification efforts; practical methods to determine the extent of FAIRness at the start of a project and define the goal of the FAIRification process (i.e., is the goal data findability and accessibility or interoperability and reusability?) in order to target investment; and simple, general guidance as opposed to complex, detailed workflows.

Fifth, Nawel van Lin presented insights from the retrospective FAIRification of the patient-empowered Duchenne Data Platform. A retrospective approach was chosen as opposed to the *de novo* approach used to build the VASCA registry because the Duchenne patient registry was already in existence and well in use. The experience demonstrated that retrospective FAIRification is a journey of many steps. Keys to success were to ensure there was a core team of FAIR data experts and that collaboration with external experts was sought. The retrospective FAIRification process involved creating a data transformation workflow and developing new software to transform the existing patient registry data (which consisted mainly of PROs) into data that were FAIR. Existing international standards for FAIRification were used to save time and money and not reinvent the wheel. That is, the same minimal CDEs, EJP-RD semantic data and metadata models for CDEs created for the VASCA project were used in the FAIRification process of the Duchenne Data Platform, which in turn helped further improve the EJP-RD semantic model. This demonstrates a sustainable, circular approach and the reusability of FAIR artefacts. The next steps in the development of the platform are to improve the robustness and sustainability of the FAIR features of the platform, link the platform to other data sources, share practical information with other patient organizations that wish to FAIRify their registries, and encourage patient organizations that do not have their own registries to join existing FAIR efforts.

Last, Mark Wilkinson, who contributed to the development of the semantic models used in the VASCA and Duchenne Data Platform projects presented above, provided a retrospective of the development of FAIR before the seminal 2016 publication [1] and shared his perspectives from the earliest

experiences with similar FAIRification projects. He reminded attendees of the foundation for FAIR being about “linking concepts, structures, and metadata to create a web of data and knowledge that could be understood and processed entirely by machines.” He reinforced that the starting point is a generic workflow [25] where the unique details for each project can be integrated. His vision for the successful implementation of FAIR involves promoting multidisciplinary collaboration to continue transforming existing data models and encourage their adoption as standards at the European level through the EJP-RD.

Summary of implementation FAIR data principles in practice

In summary, while FAIRification refers to making data machine-readable, the implementation of FAIR data principles happens in a human context. This means the FAIRification process should begin with understanding the end users and their unique context. FAIR project managers are vital for disrupting silo-mentalities and driving the change towards FAIR data practices. They also have a key role in communicating patients’ perspectives regarding data management, sharing, and protection of data. In addition, an agile approach and multidisciplinary teams are key to successful implementation. Similarly, combining efforts, establishing federations of data system networks, and reusing existing resources are key to sustainability and continued success of FAIR efforts. The developments related to machine-readable, dynamic informed consent (i.e., Data Use Ontology, ADA-M, and the Informed Consent Ontology) and data security will further support FAIRification efforts. Lastly, given that nearly 80% of a data scientist’s time is spent on getting data into a usable form, [26] which can amount to 6-months of work in a FAIRification project, there is much to be gained in terms of time and cost-savings. Case studies to demonstrate return-on-investment to make data FAIR are needed but will first require a “critical mass” of FAIR data. Activating the organizational power of patient organizations, such as through FAIR project managers, will be an important step to achieve this goal.

BUILDING A SOLID ECOSYSTEM FOR DATA SHARING

Involving stakeholders in FAIRification efforts and addressing persistent misconceptions that stakeholders may have about FAIR data principles [1]

are two key aspects of building a solid ecosystem for DMD/BMD data sharing. Clinicians, clinical researchers, regulatory authorities (eg, EMA) and organizations such as EURORDIS and EJP-RD are examples of key stakeholders in FAIR data efforts. Stakeholder perspectives were presented by Nathalie Goemans (University Hospitals Leuven, Belgium); Catherine Cohet (EMA, The Netherlands); Gulcin Gumus (EURORDIS, Spain); Daria Julkowska (EJP-RD, France); and Mary Wang (Fondazione Telethon, Italy).

Stakeholders’ perspectives

First, Nathalie Goemans described how clinicians and clinical researchers who treat patients with DMD/BMD are involved in local data collection efforts within hospitals, national efforts in collaboration with national health insurance entities, and international efforts, such as the Translational Research in Europe for the Assessment and Treatment for Neuromuscular Disorders (TREAT-NMD). Benefits of FAIR data collection for clinicians and clinical researchers include (1) improved insight into disease evolution and variability, genetic distribution factors, and adherence to standards of care, which can in turn inspire innovations in diagnosis and treatment; and (2) the ability to inform new research studies, establish new collaborations, and develop benchmarks for clinical trial results. FAIR data principles provide a framework for addressing interoperability issues that are barriers to data sharing. Other barriers from the clinician’s/clinical researcher’s perspective that need to be addressed include (1) establishing a common data set, data dictionary, and standard operating procedures for assessment, which depends on open and ongoing dialogue and achieving consensus; (2) ensuring that data collection requirements do not cost clinicians extra time or interrupt their daily workflow; and (3) addressing concerns about data protection and ownership. In particular, strategies to address reluctance to share unpublished data and authorship issues need to be developed.

Second, Catherine Cohet described that, from the regulator’s perspective, implementing FAIR data principles is one of the key steps in unlocking the potential of real-world data for public health in EU [27]. Enabling data discoverability will support efforts by EU member states, industry, and academia to improve the process of drug development and the performance of studies following market

authorization. In practice, regulators require easy access to disease registries to gain knowledge about the disease itself (e.g., incidence/prevalence, natural history following the standard of care, unmet medical needs), as well as to valid real-world evidence about drug utilization, safety, and efficacy to inform regulatory decisions. In September 2015, EMA launched a patient registry initiative that is expected to be completed around the third quarter of 2021. One of the key objectives was to promote the dialogue between regulators, companies and registry holders to understand barriers and opportunities of using disease registries. Regulators expect that data from disease registries are (1) adequate (e.g., precision of effects, range of population characteristics that are covered, length of follow-up); (2) derived from sources of demonstrated good quality; (3) accurate; (4) valid (i.e., best epidemiological and statistical practices are followed); (5) consistent (e.g., across countries or data sources) or that differences can be explained; and (6) transparent, which includes knowing how people use the registries.

Third, Gulcin Gumus presented the activities of EURORDIS, which helps shape policies related to data sharing and data protection in research and health care for rare diseases through an evidence-based approach and advocacy. Findings from a recent survey demonstrated that 95% to 97% of patients with rare diseases, regardless of the severity of their disease and their socio-demographic profile, are willing to share their data but they want to have control over the data they are sharing; in comparison, the willingness of the general population to share data ranges from 37% to 80% [28]. Furthermore, EURORDIS established so-called European Patient Advocacy Groups (ePAG) in 2019 to ensure that patients' perspectives are represented in the research and registry activities conducted by European Reference Networks (ERNs). A special working group provides support and training to ePAGs through quarterly teleconferences to discuss issues related to the quality of rare disease registries and implementing FAIR principles, and webinars to share knowledge about FAIR data resources and initiatives. The next step for ePAGs is an initiative to increase the research and innovation capacity of ERNs by encouraging inter-ERN collaboration (through ERICA – European Rare Disease research Coordination and support Action). Furthermore, EURORDIS has conducted a foresight study in which stakeholders were asked to propose policy recommendations that would lead to policy improvements and a better future for people

living with rare diseases [29]. Recommendation 7 from the study was related to optimizing data for patients and societal benefit. Key components of this recommendation included that all European data sources should be linked together as a federation; “sharing of data for care and research [is] optimized across infrastructures and countries”; data ecosystems are linked through FAIR data principles; and practices related to data sharing should reflect the “preferences and privacy of people living with rare diseases and their families.” Finally, EURORDIS launched a Rare Diseases GO-FAIR patient network in January 2021 to identify the needs of patient organizations and to be the single point of contact for developing patient-oriented FAIRification support.

Lastly, Daria Julkowska described how the EJP-RD contributes to building a solid FAIR data ecosystem through support schemes and a FAIR virtual platform of data, resources, and tools. With respect to support schemes, the EJP-RD provides funding to health care professionals, researchers, and patient advocacy organizations to establish or expand existing transnational (clinical) research networks. The EJP-RD also provides funding for research projects, research training workshops, and mobility fellowships; offers free training on, among others, data management and quality; and facilitates contact with dedicated experts in its network. The platform provides access to the following resources: experts in rare diseases, support for translational and clinical research, tools for research, registries and biobanks, animal models and cell lines, data disposition and analysis, and standards and methods. Also of interest are the findings of surveys that the EJP-RD performed as part of the virtual platform development. These findings were presented by Mary Wang. For example, survey respondents (i.e., researchers in academia and at ERNs) indicated that services which enable them to find data were more important than services which improved data management (e.g., annotating data with ontologies or standards, and data sharing), which may be reflective of their professional focus. Also, nearly half of the respondents did not sufficiently understand what FAIR research data meant and were often not aware of existing research infrastructures or resources. These findings highlight the need to demonstrate the importance of data management and to inform research stakeholders about existing infrastructures and resources so that these tools are not underutilized or unnecessarily duplicated.

Persistent misconceptions

One common misconception about FAIR is that it only facilitates humans in finding and accessing data. In actuality, FAIRification pertains to *machines* being able to read all types of data and machines knowing what humans mean by the data they create. If data are machine-readable, then machines can do the work of finding, accessing, and integrating data so that data are reusable by humans. Furthermore, contrary to what is commonly believed, the process of sharing data under FAIR principles does not entail the physical movement of data from the owner of the data. Rather, it involves sharing by means of allowing users to *visit* data virtually and conduct federated analyses without the data having to leave their original storage location. As such, owners of the data do not lose control or ownership of their data, which is a commonly reported fear behind the reluctance to share data. Moreover, FAIR data does not mean *open* data where data are freely available to the public without any safeguards. Under FAIR data principles, deidentified data are only accessible under specified conditions that conform with GDPR requirements for informed consent and data privacy and security.

It is also important to understand the difference between the quality of data and provenance of data (i.e., the record trail that accounts for the origin of a piece of data and background on how and why the data are in their present location). From the perspective of FAIR data principles, provenance of data is as important as the quality of data. The value of provenance is highlighted by the current reproducibility crisis, which is mainly due to researchers providing insufficient provenance information about methodological aspects such as workflows and versions of software to allow adequate replication. If such provenance information were available, FAIR data systems could allow users to filter data based on methods used and ask questions regarding the (methodological) quality behind the collected data.

Lastly, we should stress that FAIR is not the opposite of unfair. Similarly, striving to achieve 100% FAIRness per se may not necessarily be a constructive or relevant goal, as the focus should be on understanding the extent of FAIRification needed to achieve the practical goals of machines and people being able to reuse data. From a technological standpoint, it is always possible “to do better” but not always necessary for the situation at hand; that is, a simple

solution may suffice and highly advanced ontologies may not always have added value. Thus, in this sense, FAIR is a dynamic concept that is context dependent.

Summary of building a solid ecosystem for data sharing

FAIR benefits all stakeholders. Ongoing dialogue among stakeholders – clinicians, researchers, regulators/policy makers, companies, patients/families/patient organizations – is key to building a solid ecosystem for DMD/BMD data sharing. Through dialogue, misconceptions can be corrected and engagement and collaboration among stakeholders can be strengthened.

KEY LESSONS LEARNED

From the experiences since the first FAIR DMD/BMD data sharing meeting in 2019, eight key lessons have been identified from the presentations and discussions during the meeting. These are summarized in Box 1.

1. DMD/BMD patient organizations are key drivers of implementing FAIR in practice and the dialogue between all stakeholders is critical.
2. Creating a FAIR patient registry is a journey of many steps which begins with establishing a core FAIR team with expertise from multiple disciplines (eg, domain experts, semantic data model specialists, FAIR project manager).
3. Existing international standards should be used so that time and money are not wasted on reinventing the wheel.
4. Instead of creating central databases and moving data into them, there is a need for federated data networks that allow data to be visited and processed according to appropriate access rules.
5. There are standard questions to ask during a FAIRification process, but the answers will always be different, reflecting the unique context of each situation. As such, general guidance works better than complicated workflows when implementing FAIR data principles in practice.
6. One cannot say whether prospective or retrospective FAIRification is better compared to the other. For example, in the situation where a well-running database and application system for users to enter data already exists, making

Box 1. Summary of key lessons learned about implementing FAIR data principles in practice since the first Duchenne and Becker muscular dystrophy data sharing meeting in March 2019

- Patient organizations are key drivers of FAIRification and dialogue between all stakeholders is critical.
- The essential first step of any FAIRification process is establishing a core, multidisciplinary FAIR team.
- Use existing international standards so time and money are not wasted with reinventing the wheel.
- Federated networks with appropriate access rules, as opposed to central databases, are needed.
- FAIRification involves asking standard questions and working out the unique answers.
- While prospective FAIRification is easier than retrospective, but whether it is better depends on the project's context.
- Be sure to understand your end user's needs and context before designing and implementing any changes.
- FAIR Project Managers play a critical role in managing stakeholders' needs and expectations.

small adjustments to the existing infrastructure (retrospective FAIRification) would make more sense than starting anew. Nevertheless, starting with FAIR from the start is *easier* to do than FAIR in hindsight.

7. While FAIRification refers to making data machine-readable, the implementation of FAIR data principles happens in a human context. This means FAIRification projects need to begin by clearly understanding the type of data end users wish to collect and how, as well as what their wishes regarding data analysis and reuse are.
8. FAIR project managers employed by patient organizations have a valuable role in increasing awareness about FAIR among stakeholders. They serve as a critical link between managers of data that should be reused, database software providers, and data scientists who can make machines understand data for reuse, ensuring that expectations are defined and solutions are found.

NEXT STEPS

During the presentations and discussions, the following “next steps” regarding patient-empowered data registries were identified: (1) to encourage DMD/BMD patient organizations without patient registries to join the efforts of DMD/BMD data collection; (2) to safeguard that all patient-empowered registries can communicate with each other by sharing experiences and solutions so that they can be reused (e.g., through www.FAIRsharing.org); (3) to explore the possibility to identifying a core data set (including PROs) while recognizing limitations due to country-specific contexts; (4) to ensure that patients truly understand what they are consenting to when giving informed consent and participate in the further development and use of dynamic informed consent; (5) to ensure data entry procedures for the registries reflect the preferences of the patients; (6) to develop strategies to facilitate longitudinal data collection; (7) to continue to be aware of country-specific legal requirements for privacy and data usage while ensuring data accessibility; and (8) to further improve the interconnection of DMD/BMD data sources.

Finally, the next steps identified with respect to facilitating the implementation of FAIR data principles in the DMD/BMD data ecosystem were (1) to develop and disseminate best practices that provide guidance on how to determine the degree of FAIRness at the start of a project and define the target goal; (2) to collect case studies to demonstrate the business case and return-on investment for implementing FAIR data principles in practice; (3) to demand that software providers follow FAIR principles; (4) to establish real-world data as a source of evidence; (5) to ensure data are of known quality and representativeness; (6) to continue to advance machine-readable informed consent, generalized authentication, and authorization mechanisms; (7) to continue to engage and promote dialogue among stakeholders including patient organizations; (8) to continue to address persistent misconceptions about FAIR data, and in particular the fears regarding loss of ownership or control of data; (9) to develop strategies to facilitate scalability in practice. For example, in the clinical setting, FAIRification should be embedded in the software of data systems so that the usual workflow of clinicians is not disrupted by extra documentation (data collection) requirements; and (10) continue to increase awareness about FAIR data stewardship.

FINAL REMARKS

Since the first FAIR data sharing meeting in 2019, the DMD/BMD patient community has made considerable progress toward its goal of making DMD/BMD-related data FAIR. The DMD/BMD patient community is a pioneer among rare diseases when it comes to seeking collaborative solutions with its stakeholders. Patient organizations along with FAIR project managers are essential drivers of FAIRification and are key to ensuring that stakeholders do not lose sight of the ultimate goal of improving diagnosis, treatments, and health care policies to benefit patients with DMD/BMD.

ACKNOWLEDGMENTS

We would like to express our gratitude to all the speakers, moderator, and attendees for their enthusiastic participation in this meeting, and their support of the Duchenne FAIR data endeavor. We thank the Duchenne Data Foundation for financial support to hold this meeting and to create the FAIR animation video entitled, “FAIR for Duchenne data (what Justus wants)” (<https://www.youtube.com/watch?v=MHZ377PsHZ4>). Lastly, we thank Kimi Uegaki, PhD (iWrite) for her assistance with preparing the manuscript.

CONFLICT OF INTEREST

Nawel van Lin, Georgios Paliouras, Marco Roos, and Peter A.C. 't Hoen declare that they do not have any conflicts of interest. Elizabeth Vroom is the CEO of the Dutch Duchenne Parent Project and sponsors the Duchenne FAIR project.

REFERENCES

- [1] Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. Comment: The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*. 2016;3. doi: 10.1038/sdata.2016.18
- [2] Landi A, Thompson M, Giannuzzi V, et al. The “A” of FAIR – as open as possible, as closed as necessary. *Data Intelligence*. 2020;2(1-2):47-55. doi:10.1162/dint.a.00027
- [3] Brewster C, Nouwt B, Raaijmakers S, Verhoosel J. Ontology-based access control for FAIR data. *Data Intelligence*. 2020;2(1-2):66-77. doi:10.1162/dint.a.00029
- [4] World Duchenne Organization. FAIR data for Duchenne. Accessed May 5, 2021. <https://www.worldduchenne.org/fair-data-for-duchenne/>.
- [5] Vroom E, Isla J. Patient-managed registries: how can we get high quality data from patients for precision medicine? Accessed May 5, 2021. <https://www.ema.europa.eu/en/documents/presentation/presentation-52-patient-data-platform-e-vroom-j-isla.en.pdf>.
- [6] Verhaart IEC, 't Hoen PAC, Roos M, Vroom E. Meeting on data sharing for Duchenne 21–22 March 2019 Amsterdam, the Netherlands. *Neuromuscular Disorders*. 2019; 29(10):800-810. doi:10.1016/j.nmd.2019.08.010
- [7] Duchenne Data Foundation. FAIR data sharing for Duchenne meeting. Published March 4, 2021. Accessed July 1, 2021. <https://www.youtube.com/watch?v=SMvFrWSjg8&t=4239s>.
- [8] Centers for Medicare & Medicaid Services. Interoperability and patient access fact sheet. Accessed May 6, 2021. <https://www.cms.gov/newsroom/fact-sheets/interoperability-and-patient-access-fact-sheet>.
- [9] Office of the National Coordinator for Health Information Technology T. What is FHIR®? Accessed May 6, 2021. <http://www.hl7.org/fhir>
- [10] The Office of the National Coordinator for Health Information Technology. About ONC’s Cures Act final rule. Accessed May 6, 2021. <https://www.healthit.gov/curesrule/overview/about-oncs-cures-act-final-rule>.
- [11] Parent Project Muscular Dystrophy. PPMD & Sarepta partner to launch Duchenne outcomes research interchange. Accessed May 6, 2021. <https://www.parentprojectmd.org/ppmd-sarepta-partner-to-launch-duchenne-outcomes-research-interchange/>.
- [12] Duchenne Parent Project. Duchenne data platform. Accessed May 6, 2021. <https://duchenne.nl/duchenne-data-platform/>.
- [13] CABl.org. Accessed May 7, 2021. <https://www.cabi.org/>.
- [14] GO FAIR. Virus outbreak data network (VODAN). Accessed May 7, 2021. <https://www.go-fair.org/implementation-networks/overview/vodan/>.
- [15] The Personal Health Train. The personal health train network. Accessed May 7, 2021. <https://pht.health-ri.nl/>.
- [16] Beyan O, Choudhury A, van Soest J, et al. Distributed analytics on sensitive medical data: The personal health train. *Data Intelligence*. 2020;2(1-2):96-107. doi:10.1162/dint.a.00032
- [17] European Open Science Cloud. EOSC portal. Accessed May 7, 2021. <https://eosc-portal.eu/>.
- [18] European Commission. Digital health data and services – the European health data space. Accessed May 7, 2021. <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12663-Digital-health-data-and-services-the-European-health-data-space.en>.
- [19] FAIRsharing. Accessed May 6, 2021. <https://fairsharing.org/>.
- [20] dos Santos Vieira B, Groenen K, 't Hoen PAC, et al. Applying the FAIR data principles to the registry of vascular anomalies (VASCA). In: *Studies in Health Technology and Informatics*. Vol 271. IOS Press BV; 2020:115-116. doi:10.3233/SHTI200085
- [21] Kersloot MG, Jacobsen A, Groenen KHJ, et al. De-novo FAIRification via an electronic data capture system by automated transformation of filled electronic case report forms into machine-readable data. *medRxiv*. Published online March 8, 2021:2021.03.04.21250752. doi:10.1101/2021.03.04.21250752
- [22] Groenen KHJ, Jacobsen A, Kersloot MG, et al. The de novo FAIRification process of a registry for vascular anomalies. *medRxiv*. Published online December 14, 2020:2020.12.12.20245951. doi:10.1101/2020.12.12.20245951

- [23] VASCA common data elements (CDE) - datasets. Accessed May 8, 2021. <https://decor.nictiz.nl/art-decor/decor-data-sets-vasca-?id=&effectiveDate=&conceptId=&conceptEffectiveDate=>.
- [24] Home · ejp-rd-vp/CDE-semantic-model Wiki · GitHub. Accessed May 8, 2021. <https://github.com/ejp-rd-vp/CDE-semantic-model/wiki>.
- [25] Jacobsen A, Kaliyaperumal R, da Silva Santos LOB, et al. A generic workflow for the data FAIRification process. *Data Intelligence*. 2020;2(1-2):56-65. doi:10.1162/DINT_A_00028
- [26] Press G. Cleaning Big Data: Most time-consuming, least enjoyable data science task, survey says. *Forbes*. Published online March 23, 2016. Accessed March 21, 2021. <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/?sh=46ad5ca46f63>.
- [27] European Medicines Agency. Ten recommendations to unlock the potential of big data for public health in the EU. Published January 20, 2020. Accessed March 25, 2021. <https://www.ema.europa.eu/en/news/ten-recommendations-unlock-potential-big-data-public-health-eu>
- [28] Courbier S, Dimond R, Bros-Facer V. Share and protect our health data: an evidence based approach to rare disease patients' perspectives on data sharing and data protection - quantitative survey and recommendations. *Orphanet Journal of Rare Diseases*. 2019;14(1). doi:10.1186/s13023-019-1123-4
- [29] Kole A, Hedley V. Recommendations from the Rare 2030 Foresight Study: The Future of Rare Diseases Starts Today. 2021.