Available online at www.sciencedirect.com

ScienceDirect

Neuromuscular Disorders 29 (2019) 800–810

www.elsevier.com/locate/nmd

Workshop report

Meeting on data sharing for Duchenne 21–22 March 2019 Amsterdam, the Netherlands

Ingrid E.C. Verhaart[a,b,*], Peter A.C. 't Hoen[c,1], Marco Roos[b,1], Elizabeth Vroom[a,d], On Behalf of the Workshop Participants

[a] *Duchenne Parent Project NL, the Netherlands*

[b] *Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands*

[c] *Center for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, the Netherlands*

[d] *World Duchenne Organisation (UPPMD), the Netherlands*

Received 13 August 2019; accepted 21 August 2019

1. Introduction

Thirty-six representatives of academia, clinics, industry and patient organisations from 10 countries (Belgium, the Netherlands, Spain, Italy, Greece, France, United Kingdom, Switzerland, United States, Canada) attended the 'Meeting on data sharing for Duchenne' organised by the Duchenne Parent Project. During the meeting the current status, opportunities and challenges of data sharing in the Duchenne field were discussed.

The aim of the meeting was to optimize the (re)use of data from DMD patients, collected by different stakeholders and to create a declaration that will guide future steps towards Duchenne data being more findable, accessible, interoperable and reusable, for humans and computers (FAIR) as a new paradigm for efficient data sharing and analysis. This includes a discussion on strategies to (1) convince all parties of the necessity of better data sharing and getting them involved, (2) raise awareness of newer concepts of data sharing [1], (3) making data suitable for sharing by implementing FAIR principles [1] and (4) conceptualize ways to optimize data sharing platforms to achieve the goals 1–3.

1.1. Background

1.1.1. Duchenne muscular dystrophy

Duchenne muscular dystrophy (DMD) is a severe and progressive muscle wasting disorder with an incidence of around 1 in 5000 newborn boys [2]. It is caused by a mutation in the *DMD* gene encoding for the dystrophin protein [3,4]. Due to the absence of dystrophin, muscles easily get damaged during contractions, resulting in the gradual loss of muscle tissue [5]. Since muscle function declines over time, patients become wheelchair-bound around 12 years of age, require assisted ventilation in their late teens and develop cardiomyopathy, eventually leading to premature death, usually before the age of 30 [6,7].

1.1.2. Disease and patient data

A wealth of data on a wide number of variables related to multiple aspects of DMD exists. Broadly speaking this data can be divided into five categories:

(1) Pre-clinical data. Prior to any study initiated in a clinical setting, much data is collected in the preclinical setting to understand therapeutic mechanism of action and potential drug targets. Data is obtained through both cell and animal research prior to advancing into human testing. To date, a large range of animal models have been discovered of and generated for DMD [8].
(2) Clinical trial data. Clinical trials are conducted to collect data regarding the safety and efficacy of any new drug or therapy. These data include data reported by a clinician on patients or the patients themselves (see Patient Reported Data). Clinical trial data comprises both genotypic and phenotypic information, data on the disease progression through collection of clinical outcome measures, and other data elements included in the trial protocol. Clinical trial data reported by

* Corresponding author at: Duchenne Parent Project NL, the Netherlands.
E-mail address: ingrid.verhaart@worldduchenne.org (I.E.C. Verhaart).

[1] Both authors contributed equally to this work.

https://doi.org/10.1016/j.nmd.2019.08.010
0960-8966

clinicians is often held for long periods of time while sponsors prepare packages for regulatory submission.

(3) Patient reported data. Patients are able to report data on their condition through the use of patient (and caregiver) self-report registries. Patient reported data includes the collection of patient reported outcome measures (PROMs) which are data submitting by patients about the status of their health condition, through survey instruments, without interpretation by a clinician. PROMs can be collected for both clinical trials and through patient registries where patients report on current standard of care. This data also includes observer reported data because Duchenne is a paediatric condition and normally it is a caregiver who reports the data on behalf of their child.

(4) Natural history data. Natural history studies are used to track the course of Duchenne over time. They can be used to identify demographic, genetic, environmental and other variables that correlate with the disease and outcomes in the absence of new treatments. Natural history data informs drug development, patient care best practices, research priorities, and clinical trial readiness. A number of studies have been completed or are ongoing on the natural history course of Duchenne and changes due to improvements of care [6,7,9,10]. Large amounts of data have been gathered to date globally.

(5) Real world data. Data collected during routine clinical visits with the patient (normally entered through electronic health records). Real World Data also includes post market studies for approved therapies. Data is currently being collected on the first approved therapies in Duchenne (Emflaza and Exondys 51 in the US, Translarna in Europe). There is emerging research into the collection of data within a home-based setting such as the use of wearable devices and video capture. At the moment these are still emerging areas of research, the goal is that through new technologies families can contribute additional data on the Duchenne disease progression.

Currently these data are scattered and thereby difficult to find and access. At the moment 80% of all data generated, gets lost within two years (i.e. it is not findable and/or accessible anymore) and 33% of time and budget spent on data in the health domain is related to making bad data more usable [11]. Comparing and combining data is very hard, because many different data formats are used, and information – for humans and computers – on how to interpret data values is insufficient or missing (*e.g.* a column header is hardly sufficient). Data from academic studies and clinical trials are often made accessible only through published journal articles, with supporting data in formats that are very hard to interpret. This makes it impossible to exploit the power of computers for efficiently analysing data in larger studies across multiple sites. Furthermore, large quantities of data from the studies are often not made publicly available. A recent survey of the European Joint Program Rare Diseases (EJPRD) on the European Reference Networks revealed that a vast majority of research data managed by ERNE healthcare providers is stored locally and not made available through any global repository (Wang and Schaefer et al., manuscript in preparation). This severely hampers both scientific research and clinical improvements. Therefore, there is an urge to enable data sharing that optimizes and fully realizes the use of all types of DMD data. During the workshop the challenges and opportunities of data sharing were addressed.

2. Session 1: fair data

Elizabeth Vroom (Duchenne Parent Project, the Netherlands; World Duchenne Organization) opened the meeting stressing the need for a universal format of data and platform for sharing. Duchenne patient organizations have been working on this subject for several years and set up the Duchenne Data Foundation[1] to enhance and facilitate activities in this field.

2.1. New concept of data sharing

Barend Mons (President of CODATA, The Committee on Data for Science and Technology of the International Council for Science, former chair of the High Level Expert Group on the European Open Science Cloud, the Netherlands) discussed the need for changing the view on the definition of data sharing. One of the main hurdles for data sharing is people fearing to lose control by actively having to give their data away. To many, sharing data means collecting data from different sources and storing the data in one place. This causes several issues: the data have to be centrally harmonized and updated, privacy problems (e.g. non-compliance with the new General Data Protection Regulation (GDPR) policy of the European Union [12]), and no means to trace the original source (and thereby) quality of the data. Eventually it becomes very hard to keep covering the cost of the central store and expose the widest possible range of data types for reuse. Mons argues therefore that the concept of data sharing should be replaced by the concept of data visiting. This means that the data stay in their original place and are controlled by the person or institute who entered, generated or collected the data, but can be accessed ('visited') by third parties, which have specifically gotten permission to access a certain subset of data. Also psychologically this makes a huge difference. The key principle to enable this concept is that all data should become FAIR [1]. Implementing FAIR principles has become a global initiative. Several European initiatives, have been launched to contribute to a global ecosystem of FAIR data. The European Open Science Cloud (EOSC)[2] is an initiative, supported by the European Commission, to provide public data complying to FAIR data principles [13–16]. This requires culture changes, training and technology. GO

[1] https://www.duchennedatafoundation.org
[2] http://ec.europa.eu/research/openscience

FAIR[3] is a public-private collaboration that aims to provide an infrastructure for implementing these FAIR data principles. In the rare disease domain, a rare disease GO FAIR implementation network (RDs GO FAIR) was started to foster cultural change,[4,5] while the European Joint Program Rare Diseases (EJPRD) has adopted FAIR principles as the basis for building its infrastructure. The EJPRD works with all European Reference Networks of rare disease expert centres, EURORDIS, the European Life Science Data infrastructure ELIXIR,[6] the Biobanking infrastructure consortium BBMRI-ERIC, and the Global Alliance for Genomics and Health to achieve its goals.

Marco Roos (Leiden University Medical Centre, the Netherlands) has a leading role in FAIR activities in the rare disease community; he co-leads work in the EJPRD and the rare disease community in ELIXIR, and initiated RDs GO FAIR. Marco stated that most research questions need multiple resources to be answered; however, at the moment most data are kept in silos with no mutual communication. Therefore, gathering all data for one analysis, and making them compatible, is time-consuming and error-prone and has to be repeated for every analysis. Ideally, answering cross-resource questions takes minutes, or at most hours, on a *virtual* platform of communicating resources. There, the process of finding, accessing, interoperating, and reusing data is highly automated, because FAIR principles are implemented at each resource. Pilots by the EU-funded rare disease infrastructure project RD-Connect showed encouraging results about its feasibility [17–21]. An important aspect is the standardisation of data (*e.g.* terminology, coding) to become interpretable for machines, enabling automatization. Making data FAIR compliant requires input from data analysts, patients and clinicians. Although this is a large effort, it will save more time and costs when data sets are reused multiple times, which should be expected of valuable rare disease data. The pathway for adopting the FAIR principles is shown in Fig. 1.

Peter-Bram 't Hoen (Radboudumc, the Netherlands) introduced the concept of the Personal Health Train (PHT; Fig. 2). Data are kept in their own place in personal data lockers (stations) and do not leave this station (distributed data stations). Third parties (e.g., researchers, clinicians, commercial parties) can evaluate if the data in the locker are interesting for them and ask for access. Thereafter their queries or algorithms (trains) can be sent to the data and get access to the relevant parts for which they have permission. Combining the results from these queries or algorithms will answer a question (distributed learning). FAIR makes this process highly automated, such that no sensitive data will be seen by third parties without permission. This guarantees validation and privacy of the data. To facilitate automatization of this process, data should not be stored as text (e.g. research papers), since this is hard for machines to interpret (automatization). Validation of the quality of data is hard for a machine, but if the source is known exactly, it can decide to use or not use certain types of data. If needed, temporary umbrella stations can be formed by parties with the same interest (for a visual explanation of the PHT, see https://vimeo.com/143245835). Local pilot projects have demonstrated its feasibility, but there is still a long way to go before it can be done at a global scale. One of the hurdles for implementation of the PHT principle in DMD is that several parties are involved. Different stakeholders (*e.g.* patients, clinicians, researchers, industry, regulators and payers) have different incentives to participate. Their collaboration should be improved, for example via the RDs GO FAIR network. Another issue is that the current legislation is outdated and does not comply with the current situation. Duchenne parent project (DPP) organisations can play a role in designing a new legal and ethical framework compatible with the GDPR. The personal data lockers enable both sharing and compliance with the GDPR. Every patient has access to his/her own data, can manage it and give dynamic consent that he/she can change if desired. The need for and type of consent depends on the kind of data (sensitive or non-sensitive) and the party asking for it (*e.g.* academic or industry). The PHT also has important advantages for the patients themselves. This system provides a way for patients to receive feedback on their data. Secondly, patients can find patients with similar characteristics (*e.g.* same mutation). This enables them to compare their own health condition to others or to come into contact with other patients. The infrastructure required is, however, not yet in place.

2.2. FAIRification of data from different stakeholders' perspectives

We use the term 'FAIRification' for the process of making a data source FAIR (Fig. 1).[7] FAIRer for humans, for instance, could pertain to registering a specific Duchenne registry in the European rare disease platform, a 'registry of registries' set up by the European Joint Research Council to list all EU rare disease registries.[8] Usually, data are already reasonably FAIR for humans (*e.g.* through a web site), such that most effort goes into making data FAIR for computers. It is where the largest efficiency gain comes from (up to saving months of work per reuse of a data set). It entails recoding data and access permissions into global, machine readable standards, and complementing local data management solutions with FAIR data access tools. Many different stakeholders are involved in the FAIRification process, each having their

[3] https://www.go-fair.org/
[4] https://www.go-fair.org/implementation-networks/overview/rare-diseases/
[5] https://tinyurl.com/rdsgofairform
[6] https://elixir-europe.org/

[7] For a workflow and an overview of considerations regarding the steps to make a resource FAIR we refer to 'A generic workflow for the data FAIRification process', by Jacobsen et al. (in 'FAIR in practice: First Generation implementation choices and challenges', Jacobsen, Schultes, Mons, Eds., Data Intelligence Journal special issue, submitted), which builds on earlier work for the rare disease domain (https://zenodo.org/record/1452468#.XUqgnOgzYis)
[8] https://eu-rd-platform.jrc.ec.europa.eu/erdri-description

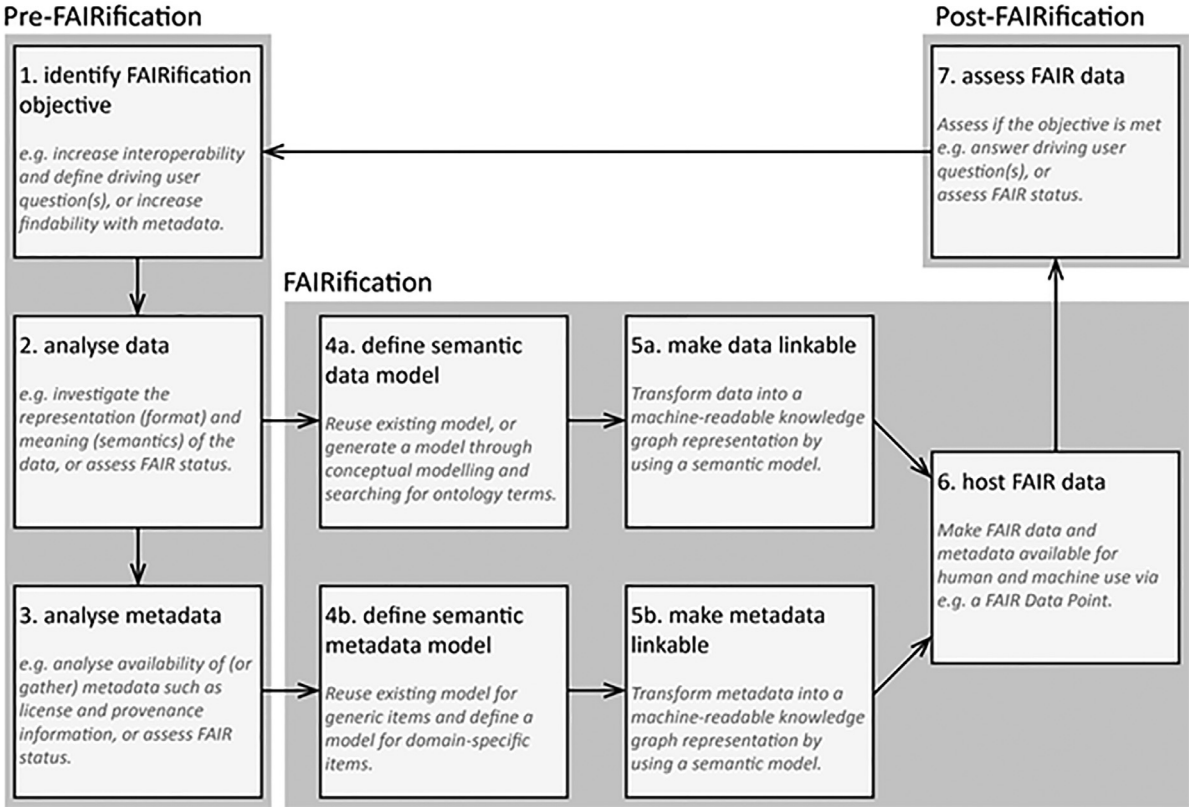

Fig. 1. Pathway for implementing FAIR principles (reused with permission from Fig. 1 in Jacobsen et al., 2019 (FAIR in practice: First Generation implementation choices and challenges, Jacobsen, Schultes, Mons, Eds., Data Intelligence Journal special issue, submitted)). A generic step-by-step workflow for the process of making data FAIR ('FAIRification'). The workflow is divided into three 'phases': Pre-FAIRification, FAIRification, and Post-FAIRification (dark grey boxes) that are further specified by 'steps' indicating typical aspects of practical FAIRification (light grey boxes): (1) identify FAIRification objective, (2) analyse data, (3) analyse metadata, (4a) define semantic data model, (4b) define semantic metadata model, (5a) make data linkable, (5b) make metadata linkable, (6) host FAIR data, and (7) assess FAIR data. The order is not strict and can be iterative.

own incentives, benefits, opportunities and challenges to contribute.

Patient representatives can advocate the new level of openness and reusability of data that is achieved by implementing FAIR principles (*e.g.* by implementing 'accessible under well-defined conditions' for computers). They can inform patients and other stakeholders about the importance of sharing data in a universal, accessible form and guide them in their own attempts to stimulate the FAIRification of data (*e.g.* by organising training meetings). The Rare Diseases GO FAIR implementation network offers to provide a platform to foster such attempts. User-friendly, 'FAIR data generating' software is an absolute requirement to scale up FAIRification, but it may not be advisable to entirely surpass interaction between human experts, given the importance of correctly representing the meaning and access permissions of valuable and sensitive rare disease data. In this context, efforts are needed to reach and include the less empowered patients and parents. Patient representatives can also advocate new regulations for sharing medical data. They can fund projects aiming at FAIRifying data. Additionally, when they grant projects that generate or manage data, they can include a requirement for FAIR data stewardship and reserving budget for FAIRification. At this time, requiring detailed technical FAIRification plans upfront is not advisable; budget for a FAIRification team is much more important. The RDs GO FAIR network can play a role in finding partners with FAIRification experience.

A review is needed of what kinds of data are already available and how to make the best use of them. Patient representatives can contribute to gathering more patient-orientated data to increase the potential for addressing patient priorities, such as research towards improving individual care and Quality of Life (QoL). For the patients/parents themselves it provides a way to share how they/their son is doing and compare this to others with a similar mutation. This gives them strength and control. It is worth noting that FAIR electronic data are more prepared for yet unforeseen future applications. The anticipation is that increased amounts of FAIR data will stimulate creative use of data by all stakeholders.

Gathering FAIR data is most efficient when done from the start of a project and discussions should be held with industry to make sure they do the same. Important aspects are the use of common computer languages (harmonisation) across centres, implementation of quality systems and making sure the consents contain all required elements. Therefore, discussions are needed at (inter)national levels, but also

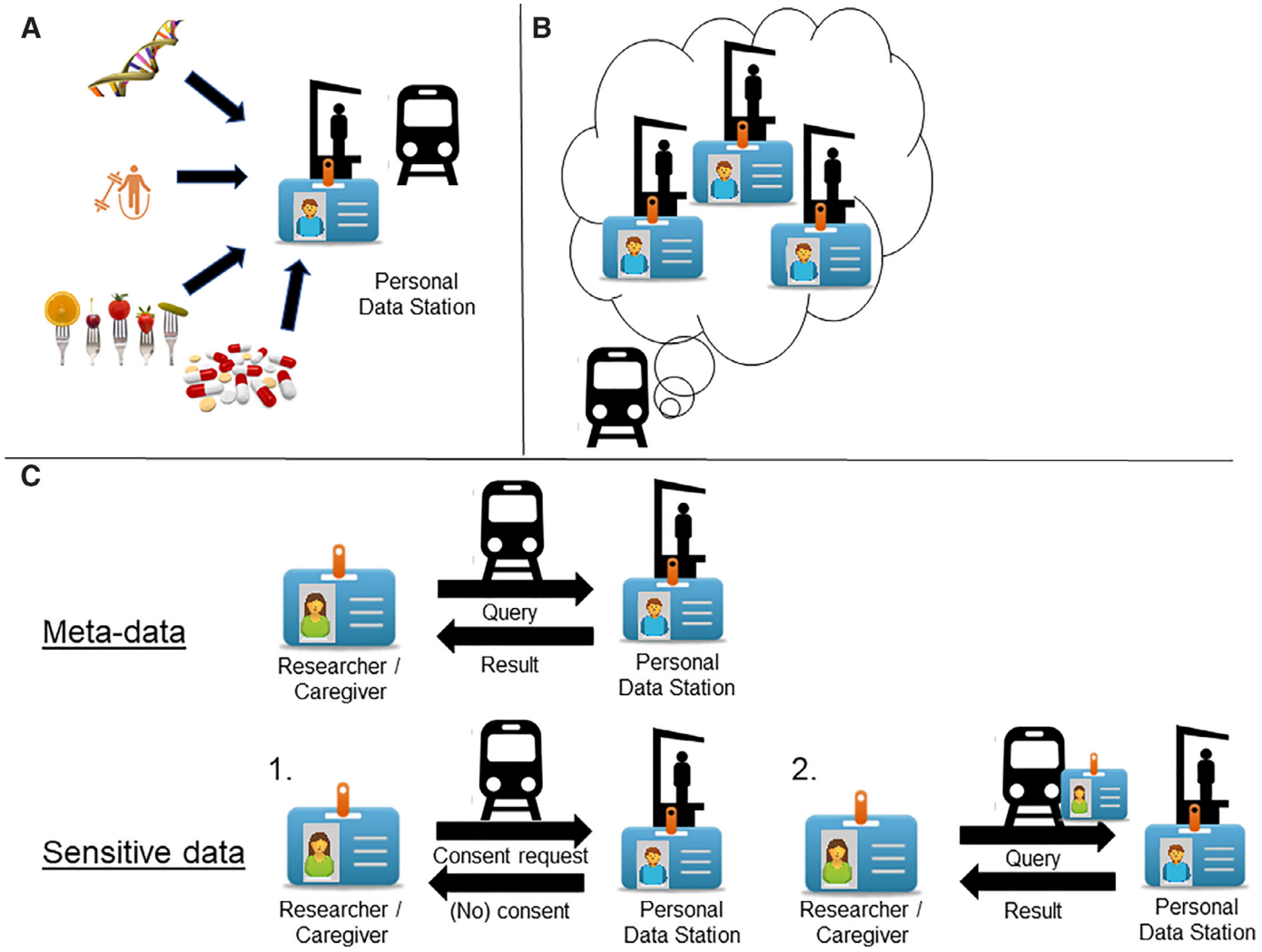

Fig. 2. The Personal Health Train perspective on personal data sharing. (A) Different types of personal data are managed from a personal data station. The station contains FAIR data that can be visited by trains carrying the queries or algorithms to be applied to the personal data. The data may not always be fully contained within the locker, but may also contain pointers to the data (for example hospital data). (B) Patients and patient organizations may organize themselves through building of platforms containing multiple personal data stations, which can be accessed simultaneously. (C) Data do not leave the station. Third parties (*e.g.* researchers or caregivers) can ask for access. Meta-data can be queried from personal data stations (A) or joint stations (B). Identifiable (sensitive) data can only be queried after consent given by the patient or patient organization.

within hospitals (*e.g.* with the IT department). A possible barrier for scientists is that they are still solely judged on narrative text publications and citations thereof. Credit for other types of contribution, such as for data and collaboration is still in its infancy. Therefore, it should be made an obligation to publish (*i.e.* FAIRify) all data from publicly funded projects, next to traditional textual publication, with an embargo time if needed. Many European funders already do not fund projects anymore if data will not comply with the FAIR principles [1,22]. For researchers/clinicians themselves, FAIR data will help to improve the development of treatments. Extensive natural history data of high quality will enable assessing drug effects more quickly and more comprehensively, also outside a clinical trial setting. Furthermore, the PHT also makes feedback on the consequences of a certain interventions possible. For example, after the discovery that the benefit/harm of a certain medicine depends on the patient's genotype, hospitals can adapt their policy.

Industry owns a lot of clinical trial and legislative data. A lot of time and effort is spent on collecting all data, but less on making them reusable. At the moment, efforts are made to standardise the way the data are *presented* to the regulators; however, the data themselves are not standardised to the extent of enabling PHT-like scenarios. Regulators feel that industry still has a long way to go. It should really be made mandatory by regulators. Industry could hugely benefit from FAIR data [23]. An example is the repurposing of medicines for other diseases. The IMI FAIRplus[9] and EHDEN[10] projects are promising signals that also Pharma is embracing FAIR principles.

[9] https://fairplus-project.eu/
[10] https://www.ehden.eu/

For regulators FAIR data are needed. Objective data could prevent failure of registration. Patient representatives should give incentives and advice to regulators. They can formulate guidelines and give consultations to the regulators. Setting up post-marketing FAIR registries should become obligatory, because they are necessary for the continuum of data. Thereby the effectiveness of drugs and risks/benefits can be monitored.

2.3. Discussion

Several questions and points for discussion were raised during the talks. The feasibility of being able to answer all questions and draw conclusions using distributed data was questioned. Data experts believe that making data interoperable (harmonisation) will facilitate this; however, theoretically there may be a subset of questions that cannot be answered in this manner. At the moment some deep learning algorithms are still difficult, but work is ongoing to improve this. Even if it is theoretically possible, in some cases the time it takes to answer complex questions over distributed resources may be too long (performance issue). This may be a reason to temporarily centralise data. Fortunately, 'FAIR at source' also prepares data for this: the cost of integrating data into one place is substantially reduced when data are FAIR. Consequently, such aggregations can be made transiently, which has important benefits for personal data.

Another aspect is the protection of the privacy of the data owner. The PHT concept mitigates the risk of exposing sensitive data to unauthorised people by keeping data in their own private space, allowing only authorized computer algorithms (not people) to visit the data and only non-sensitive or de-identified data to leave the station. Nevertheless, the theoretical safeguards should still be tested rigorously, in particular when combinations of data items are requested or analysed. Therefore, automated systems to detect potentially identifiable information are required.

Another hurdle to overcome is that, in the past, consents of thousands of patients have been collected for specific purposes, but not for general reuse. It is a legal and practical challenge to go back to the patients and get consent for new purposes. This can be mitigated by automating dynamic consent procedures (*e.g.* via mobile phone apps), which in turn requires that also consent is standardised and machine readable. This is a topic of active research in for instance the EJPRD project, IRDiRC and ELIXIR. Discussions include introducing multiple levels of consent of what can be done with which data under which circumstances (*e.g.* making the allowed level of identifiability depend on who is using the data). Introduction of dynamic consent supported by digital ways of communication with the patient may become an alternative to a broad, overall consent (see the THREAD platform for an example). It introduces transparency and insight in who does what with a person's data.

Financial aspects were also addressed. Getting money from the industry for making data FAIR is difficult, since the incentive/reward lies in the future. A strong case should be made for industry, clearly explaining what the extra value of a data platform is and that it will in the end reduce their costs. Clinicians/researchers should incorporate time and money for making data FAIR in the budget of their grants. It is advised to dedicate around 5% of the budget of data generating projects on FAIR data stewardship and bioinformatics.

3. Session 2: data collection for DMD

In the last years some progress has been made in developing new ways to improve data collection of DMD patients.

3.1. New initiatives in Duchenne data collection

The Duchenne field can learn from successes and failures in other disease fields, regarding many topics related to data sharing. Julián Isla (Dravet Syndrome European Federation; Foundation29, Spain) reviewed several projects in which the Dravet Syndrome European Federation (Dravet syndrome is a rare, genetic, epileptic encephalopathy [24]) has been involved (*e.g.* setting up a registry, technology and research). All these projects were in collaboration with many different partners. There have been some successes. One of their important achievements is getting an orphan drug status for the treatment of Dravet syndrome. Therefore, they actively participated as EMA committee members in clinical trial design (i.e. giving advice to companies/clinicians). The design is very important to prove effectiveness of the drug and prevent failure. An example is that one of the companies did not want to include the existing drug, already on the market, in their trial. One of the questions asked by the European Medicines Agency (EMA) was how many Dravet patients were using that drug in Europe; however, no numbers were available. The federation held a survey amongst patients. Thereafter the EMA recommended the company to include the drug in their trial. Unfortunately, most other projects failed, amongst others because it was difficult to get everybody on the same wavelength and willing to share their data in the same manner. Important lessons can be learned from these failures. More awareness is required since the need for data is not clear to everybody. It should be emphasised that missing data cost both lives and money. Currently drug development is slow. Over the past years no progress has been made in increasing the number of approved orphan drugs per year. However, if there would be drugs for all orphan diseases developed in the current way, the costs would be prohibitive. Therefore, the process should be accelerated, for which correct data capturing (i.e. machine readable) and sharing is invaluable. Foundation 29 is exploring how to empower citizens with the data they own. Health29 is the platform that Foundation 29 is creating, providing a personal health record to the patient. This personal health record (PHR) will be built following FAIR principles and it will allow patients to store data in a secured container for controlled visiting. The use of these PHR has been studied in medical literature and the patients can see the benefit of it [25]. The use of individual PHR and having connection platforms will

allow to create an Internet of Patients (IoP[11]) in the same way that Internet of Things (IoT) is being developed.

Pat Furlong (Parent Project Muscular Dystrophy (PPMD), United States) discussed the changing environment of the US DMD registry. The Duchenne registry is a self-reporting registry, which has evolved over time, but focuses on four main areas: (1) knowledge on the DMD community, (2) patient resources, (3) drug development and (4) information sharing. PPMD has also set out to enhance the Duchenne Registry itself, launching a new registry platform (partnership with THREAD Research) based on modern technology. The new technology allows patients and caregivers to give informed consent to access the electronical health records (EHR) from their site of care. Advantages are that the data can be gathered automatically and it does not rely on clinicians entering the data into another database, the data is pulled from the electronic health record of the hospital. PPMD also has a partnership with the Critical Path Institute to form The Duchenne Regulatory Science Consortium (D-RSC).[12] The Duchenne Regulatory Science Consortium (D-RSC). The D-RSC database currently includes twelve integrated databases. The D-RSC data platform will allow C-Path, members and possibly outside groups to analyse some or all of the integrated data for multiple purposes

Nathalie Goemans (KU Leuven; TREAT-NMD Global Database Oversight Committee (TGDOC), Belgium) spoke about Translational Research in Europe for the Assessment and Treatment for Neuromuscular Disorders (TREAT-NMD) network.[13] This network started as an EU funded Network of Excellence, but is now a global organisation, called TREAT-NMD Alliance. It has several resources for industry, clinicians and scientists aiming to facilitate the translation of preclinical therapeutic developments into the clinic and establishing high level standards of care for neuromuscular disorders. One of their resources are, national or regional, patient registries for multiple neuromuscular disorders (including DMD, SMA, DM1, FSHD) that have clear, parsimonious datasets developed to encourage global data harmonization for translational research [26,27]. The TREAT-NMD Global Database Oversight Committee (TGDOC) reviews requests for data from academic or industry parties and coordinated the response from the global registries. The format of the global registries is, however, outdated. After a request has been approved, the questions are sent out to the national/regional registries. These respond by sending their data, which is then combined. For SMA a platform project has been started to improve the data collection (*e.g.* natural history data and post-marketing data concerning the recently approved drug for SMA) and sharing [28]. A centralised Universal Registries Platform (URP), using Open-App software, has been set-up. Individual registries retain the ownership of the data and only de-identified data are shared with third parties. Different modules are possible and can be chosen by the individual registries. They can opt to enter data directly onto the common data platform, having a "substudy" for the registry on the platform. They can also install the SMA OpenApp module on a local/national server facilitating technically the pooling and sharing of data for analysis. The main goal is to facilitate harmonized data collection and sharing by providing a common dataset and by developing IT solutions that would enable this. Data are collected and uploaded continuously. Analysis can be done once request has been approved by TGDOC. This could serve as an example for the DMD global registry.

Yolanda Ludeña (Foundation29, Spain) introduced the Duchenne Data Platform (DDP). The DDP has been created by Foundation29,[14] a foundation aiming to use artificial intelligence to combine technology and health, and by DPP, the Netherlands. It is a user-friendly, GDPR-compliant, online environment, accessible either via internet or via an app. The development started from the principle that patients are the owners of their data. What makes this platform unique is that patients have their own 'data lockers' where they can not only collect their PROM's but can also upload their data from others sources such as hospitals and wearables. Thanks to the GDPR (portability of data) patient should have the possibility to receive their own data in a portable and machine-readable format. The lockers allow for data visiting. Since a dynamic informed consent is used, they can decide who to give access to their (anonymised) data. The patients can also ask questions to the platform themselves. At the moment only patients (or their parent/caregiver) and administrators are participating in DDP, but in the future also clinicians and researchers will be involved. It uses standard codifications to store information, such as symptoms or genetic data, in line with the FAIR principles. All data stored in cloud will be compliant with the Fast Healthcare Interoperability Resources (FHIR) standard. FHIR is rapidly gaining support in the healthcare community as a candidate for the next generation standards framework for interoperability of health data. Projects such as the EJPRD will critically appraise such frameworks and how they can interoperate. The aim for DDP is to simplify the implementation and interoperability of health data by adhering to the FHIR standard. This will enable the use of a wealth of data, which would otherwise be lost. One of the most important features to work on is 'privacy by design'. FHIR provides handles for making data highly secure and incorporating de-identification techniques, following the GDPR mandates.

3.2. Discussion

So far most of these data initiatives are regionally focused. In the end worldwide actions are needed. One thing that came out of discussions between Europe and the United States are the cultural and legal differences between countries. A

[11] https://ec.europa.eu/futurium/en/content/future-health-care-deep-data-smart-sensors-virtual-patients-and-internet-humans

[12] https://c-path.org/programs/d-rsc/

[13] https://treat-nmd.org/

[14] https://www.foundation29.org/

hospital in the US offered to sell their data, while in Europe it is not allowed to make money out of patient data. Ideally, these kinds of hurdles should be overcome by harmonising the processes in a way that allows all data resources to comply with their national laws. The question remains if this will be feasible.

One of the concerns regarding this type of platforms is misuse of data and entering of fake data. To avoid this, it is crucial that the exact source of the data is known. The data should also comprise information on the personal qualification of the person entering the data, if the data are gathered pro- or retrospectively, when the data were collected etc. At the moment this information is often lacking. Since data validation is very difficult for a machine, the investigator him-or herself should decide whether or not to include the data in the analysis.

While most of these initiatives strive to achieve high levels of FAIRness, the extent to which interoperability between the different solutions has been achieved, remains to be assessed. In other words: to what extent have the same, or at least compatible standards been used to make data FAIR and machine readable? Registry software providers may offer features for data to be communicated between installations of their own software and they usually provide download options in various formats. However, that is still far from what is possible, and far from the efficiency gain that the FAIR principles and the PHT concept are intended for. It may be tempting to think that using the same tool in the Duchenne community is the answer to interoperability. The varied list of tools presented in this meeting is an indication of how undesirable and difficult that will be. Instead, the more feasible and flexible approach is to work together on adding minimal features to these tools to *also* provide FAIR and PHT functionality, and to work with larger international initiatives, such as GO FAIR and the EJPRD, to ensure compatibility with some of the most used FAIR approaches.

4. Session 3: blockers and unblockers for FAIRification of data

Several challenges and hurdles exist for making data FAIR. On the other hand, there are opportunities that facilitate the process.

4.1. Blockers

Overall, the most importance issue is the lack of awareness at all levels. The value of data sharing (or data visiting) is underestimated. It requires effort now, while the benefits are mostly on the long run. Time and costs play a big role. The effort of preparing data for analysis is moved from data users to data producers. This ultimately saves costs overall, but this is not always visible to the data producer. It is a complex process to convince all parties of the necessity and getting them on the same page.

The current attitude of legal departments hampers data sharing. Companies, but also scientists, try to protect their data. This plays a role if you want to combine data from different companies/sources. Different views on ownership of Intellectual Property (IP) are an issue too. Companies consider the data theirs, because they initiated the trial and feel that they need free access to the data for marketing authorisation application. On the other hand, clinicians and their payers (trial site hospitals) also regard it as theirs. The general feeling is that who owns the data has the power, so sharing means a perceived loss of power.

Contrary, sharing too much information is sometimes a problem. Occasionally too much (*i.e.* patient identifiable data) is disclosed, which is not compliant with GDPR and not in the interest of individuals when it could be avoided. This makes patients more reluctant to share their data. Institutional GDPR officers often lack resources and are overwhelmed by all information. They should get independent, trustworthy advice.

Related to the previous points is the lack of (internal) trust, both between different stakeholders, but also within organisations. People feel that sharing their data with other parties will lead to monetisation of the data by the other side. The patient's/clinician's perceptive is that they put money in gathering the data. They give their data to a company, whereafter they have to pay for the medicine. They also feel it will lead to a lack of recognition of the work of the people collecting the data, which is very important for academics (*e.g.* for obtaining grants). On the other hand, industry feels that they pay clinicians/institutions to collect the data (financing clinical trials), but if they thereafter want to do something with the data (*e.g.* analyses) they have to pay again (*e.g.* researchers).

At the moment IT is a problem. There is too little expertise in hospitals/health systems. This requires education and institutions should be convinced to put money into IT departments. Platforms that offer all required functionalities for truly efficient analysis on shared data are still missing. Easy and user-friendly registry systems that require little time and effort to create FAIR data, are absolute necessary for success. Furthermore, making currently existing data (not only future data) FAIR to enable inclusion in, or interoperability with, the personal lockers, is challenging and requires a good IT-system. This also accounts for making previously obtained consents compatible with the new situation.

4.2. Unblockers

As a lack of awareness is a blocker, raising awareness will serve as an unblocker. Firstly, it should be investigated what kind of deliverables people are looking for when they share their data. Patient organisations/representatives can start campaigns to inform patients on the benefits and get them engaged. Importantly, they could inform patients to not participate in trials not adhering to FAIR data principles. This will provide an incentive for industry.

GDPR may not only be a hurdle, but can also be a solution. The purpose of the GDPR is not to block data sharing, but to ensure transparency and that the ownership of the data lies

with the patient him- or herself. It creates an incentive for IT specialists to work on efficient digital dynamic consent procedures. Patient representatives can incentivise this further by pointing at this aspect of the GDPR.

Facilitating and providing incentives would also be a stimulant for parties to get involved. Patient organisations can subsidise specific projects to include FAIRification of data to improve data sharing. In the US, patient organisations are trying to include also non-engaged patients in the registry by paying clinicians to contact them and help to keep their data up-to-date. Translation into several languages also helps.

FAIRness of data has to be stimulated so it becomes a standard best practise. Including a plan and budget for FAIR data should be made obligatory in grants, accounting for all data generated during the project. This should be checked afterwards. Journals should also make it mandatory to make source data publicly available adhering to FAIR principles.

The EMA and the European Union can also serve as unblockers by stimulating FAIRness and including it in the EMA marketing assessment procedure. The European Commission is already involved in several FAIR data projects [13,14,16,29].

5. Duchenne fair declaration

(1) Patient derived or provided data are not owned by those who collect them, and their reuse should be primarily controlled by the donors of these data. Researchers, charities, companies and health professionals are custodians
(2) To enable the optimal reuse of data, the data needs to be Findable, Accessible, Interoperable and Reusable (*i.e.* FAIR) by medical professionals, patients and in particular also by machines.
(3) The optimal reuse of data should be supported at all levels, by professionals and custodians (allow federated learning on the data upon request, give the data to the donor in FAIR format when asked), care professionals (capture data at the source in FAIR format wherever possible), analytics environments (adapt to FAIR data) and regulators (demand FAIR data throughout and optimally use them in the regulatory process). There is a need to educate all stakeholders about the FAIR principles and their importance ('FAIR Aware').
(4) Optimal care should be taken to restrict the need to reveal the actual identity of individuals associated with certain data, and to protect privacy with all possible means, but we realize privacy is subordinate in many cases to fast-tracking of better solutions for the diseases we suffer from.
(5) Therefore, the right to allow identification of the individual associated with certain data should also be placed in the hands of that individual or a chosen trusted party.
(6) Techniques and tools should be developed to enable optimal co-investigation by researchers, medical and health care professionals, charities, companies, patients and machines to form a 'social health machine' aimed at better solutions and care.
(7) Regulators should optimally enable fast-tracking of key interventions and involve citizen and machine participation in that process to the largest possible extent.
(8) The field should actively discourage publishing of health-related information exclusively in classical narrative journals. These are very difficult to access and understand by both informed lay people and by machines. Instead, data and information should be published in a way that makes it more readily reusable by others than a small inner circle.
(9) Funding agencies should have good data stewardship following the FAIR principles included in their grant conditions.
(10) The role of health insurance institutions/companies should also be made clear: they should publicly state that having full access to real world data, even when these are re-identifiable for them to 'their' clients, will not be abused, such as for example increase premiums based on genetic predisposition.
(11) Health insurance institutions/companies and governments should join forces and support (also financially) the development of trusted environments where real world citizen data can be maximally reused for the betterment of health care and the massive saving of costs to keep optimal healthcare affordable for all.

6. Workshop participants

6.1. Data experts

- Barend Mons, Department of Human Genetics, Leiden University Medical Centre, Leiden, the Netherlands
- Marco Roos, Department of Human Genetics, Leiden University Medical Centre, Leiden, the Netherlands
- Peter A.C. 't Hoen, Centre for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Centre, Nijmegen, the Netherlands Bioinformatics at Radboud UMC, Nijmegen, the Netherlands
- Julian Isla, Duchenne Data Platform; Foundation 29, Madrid, Spain
- Yolanda Ludena, Foundation 29, Madrid, Spain

6.2. Clinicians

- Nathalie Goemans, Department of Pediatrics and Child Neurology, University Hospitals Leuven, Leuven, Belgium
- Craig Campbell, London Health Sciences Centre London, Ontario, Canada
- Erik Niks, Department of Neurology, Leiden University Medical Centre, Leiden, the Netherlands
- Michela Guglieri, The John Walton Muscular, Dystrophy Research Centre and MRC Centre for Neuromuscular

Diseases, Institute of Genetic Medicine, Newcastle upon Tyne, United Kingdom
- Teresinha Evangelista Institut de Myologie, CHU Pitié-Salpêtrière, Paris, France

6.3. Patient representatives

- Pat Furlong, Parent Project Muscular Dystrophy, United States; World Duchenne Organisation
- Marisol Montolio, Duchenne Parent Project Spain, Spain
- Gary Fegan, Action Duchenne, London, United Kingdom
- Alex Johnson, Duchenne UK; Joining Jack, United Kingdom; World Duchenne Organisation
- Ilaria Zito, Parent Project Onlus, Rome, Italy
- Dimitrios Athanasiou, World Duchenne Organisation; MDA Hellas, Greece
- Mirjam Franken, Duchenne Parent Project NL, the Netherlands
- Floor Knuistingh Neven, Duchenne Parent Project NL, the Netherlands
- Elizabeth Vroom, Duchenne Parent Project NL, the Netherlands; World Duchenne Organisation
- Arjen Bergsma, Duchenne Parent Project NL, the Netherlands
- François Lamy AFM-Téléthon, France
- Sally Hofmeister, World Duchenne Organisation
- Gulcin Gumus, Eurordis, Spain

6.4. Other

- Violeta Stoyanova, European Medicines Agency, Amsterdam; Medicines Evaluation Board, Utrecht, the Netherlands
- Pietro Spitali, Department of Human Genetics, Leiden University Medical Centre, Leiden, the Netherlands
- Yvonne Meijer-Krom, Department of Neurology, Leiden University Medical Centre, Leiden, the Netherlands
- Josie Godfrey Duchenne UK, United Kingdom
- Begonya Nafria Escalera, Share4Rare; Sant Joan de Déu Hospital Barcelona, Barcelona, Spain
- George Tournavitis, Share4Rare, Omada, Spain
- Suzie-Ann Bakker, World Duchenne Organisation
- Ingrid Verhaart, Duchenne Parent Project NL; Department of Human Genetics, Leiden University Medical Centre, Leiden, the Netherlands
- Giorgios Paliouras, Duchenne Data Foundation; NCSR Dimokritos; MDA Hellas, Athens, Greece

6.5. Industry

- Abdallah Delage, PTC Therapeutics, Steinhausen, Switzerland
- Marianne Berrens, Santhera, Pratteln, Switzerland
- Matt Anderton, Wave Life Sciences, Oxford, United Kingdom
- Annika Bergman, Sarepta Therapeutics, Zürich, Switzerland

Acknowledgments

We thank the World Duchenne Organisation and Duchenne Parent Project the Netherlands for taking the initiative for and organisation of the workshop. We also thank all speakers for the critical review of this article.

References

[1] Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The fair guiding principles for scientific data management and stewardship. Sci Data 2016;3:160018.
[2] Mendell JR, Lloyd-Puryear M. Report of MDA muscle disease symposium on newborn screening for Duchenne muscular dystrophy. Muscle Nerve 2013;48:21–6.
[3] Hoffman EP, Brown RH Jr, Kunkel LM. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. Cell 1987;51:919–28.
[4] Koenig M, Monaco AP, Kunkel LM. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. Cell 1988;53:219–28.
[5] Emery AE. The muscular dystrophies. Lancet 2002;359:687–95.
[6] Eagle M, Baudouin SV, Chandler C, Giddings DR, Bullock R, Bushby K. Survival in Duchenne muscular dystrophy: improvements in life expectancy since 1967 and the impact of home nocturnal ventilation. Neuromusc Disord 2002;12:926–9.
[7] Ishikawa Y, Miura T, Ishikawa Y, et al. Duchenne muscular dystrophy: survival by cardio-respiratory interventions. Neuromuscul Disord 2011;21:47–51.
[8] Wells DJ. Tracking progress: an update on animal models for Duchenne muscular dystrophy. Dis Model Mech 2018;11.
[9] McDonald CM, Henricson EK, Abresch RT, et al. Long-term effects of glucocorticoids on function, quality of life, and survival in patients with duchenne muscular dystrophy: a prospective cohort study. Lancet 2018;391:451–61.
[10] Passamano L, Taglia A, Palladino A, et al. Improvement of survival in duchenne muscular dystrophy: retrospective analysis of 835 patients. Acta Myolog Myopath Cardiomyop Off J Mediterr Soc Myol 2012;31:121–5.
[11] Directorate-General for Research and Innovation (European Commission). PwC EU services, *cost of not having fair research data; Cost-benefit analysis for fair research data*, policy development and coordination, Luxembourg: Publications Office of the European Union; 2019. Editor.
[12] European Commission. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (General data protection regulation)* the european parliament and the council of the European Union. Official Journal of the European Union; 2016. Editor p. 1–88.
[13] Directorate-General for Research and Innovation (European Commission). *Turning fair into reality: Final report and action plan from the European Commission expert group on fair data*, open innovation and open science, Luxembourg: Publications Office of the European Union; 2018. Editor.
[14] Directorate-General for Research and Innovation (European Commission). *Prompting an eosc in practice: Final report and recommendations of the commission 2nd high level expert group on the European Open Science Cloud (EOSC)*, open innovation and open science, Luxembourg: Publications Office of the European Union; 2018. Editor.
[15] European Open Science Cloud. *EOSC declaration: European open science cloud, new research & innovation opportunities*. 2017. European Commission: Brussels.
[16] Directorate-General for Research and Innovation (European Commission). *Realising the European open science cloud*, policy

development and coordination, Luxembourg: Publications Office of the European Union; 2016. Editor.
[17] Lochmuller H, Badowska DM, Thompson R, et al. RD-Connect, neuromics and EURenOmics: collaborative European initiative for rare diseases. Eur J Hum Genet EJHG 2018;26:778–85.
[18] Sernadela P, Gonzalez-Castro L, Carta C, et al. Linked registries: connecting rare diseases patient registries through a semantic web layer. BioMed Res Int 2017;2017:8327980.
[19] Gainotti S, Torreri P, Wang CM, et al. The RD-Connect Registry & Biobank Finder: a tool for sharing aggregated data and metadata among rare disease researchers. Eur J Hum Genet 2018;26:631–43.
[20] Monaco L, Crimi M, Wang CM. The challenge for a european network of biobanks for rare diseases taken up by RD-Connect. Pathobiol J Immunopathol Mol Cellul Biol 2014;81:231–6.
[21] Thompson R, Johnston L, Taruscio D, et al. RD-Connect: an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. J Gen Intern Med 2014;29(Suppl 3):S780–7.
[22] Mons B, Neylon C, Velterop J, Dumontier M, da Silva Santos LOB, Wilkinson MD. Cloudy, increasingly FAIR; revisiting the fair data guiding principles for the European Open Science Cloud. Inf Serv Use 2017;37:49–56.
[23] Wise J, de Barron AG, Splendiani A, et al. Implementation and relevance of fair data principles in biopharmaceutical R&D. Drug Discov Today 2019;24:933–8.
[24] Dravet C. The core Dravet syndrome phenotype. Epilepsia 2011;52(Suppl 2):3–9.
[25] Patel VN, Abramson E, Edwards AM, Cheung MA, Dhopeshwarkar RV, Kaushal R. Consumer attitudes toward personal health records in a beacon community. Am J Manag Care 2011;17:e104–20.
[26] Bladen CL, Rafferty K, Straub V, et al. The treat-NMD Duchenne muscular dystrophy registries: conception, design, and utilization by industry and academia. Hum Mutat 2013;34:1449–57.
[27] Sarkozy A, Bushby K, Beroud C, Lochmuller H. 157th ENMC international workshop: patient registries for rare, inherited muscular disorders 25–27 January 2008 Naarden, the Netherlands. Neuromuscul Disord 2008;18:997–1001.
[28] Pechmann A, Konig K, Bernert G, et al. SMArtCARE - A platform to collect real-life outcome data of patients with spinal muscular atrophy. Orphanet J Rare Dis 2019;14:18.
[29] European Commission. *Commission staff working document: implementation roadmap for the European Open Science Cloud*. 2018. European Commission: Brussels.