The Multiple Sclerosis Data Alliance Catalogue

Enabling Web-Based Discovery of Metadata from Real-World Multiple Sclerosis Data Sources

Lotte Geys, PhD; Tina Parciak, MSc; Ashkan Pirmani, MSc; Robert McBurney, PhD; Hollie Schmidt, MSc; Tanja Malbaša, MSc; Tjalf Ziemssen, MD; Arnfin Bergmann, MD, PhD; Juan I. Rojas, MD, MSc; Edgardo Cristiano, MD; Juan Antonio García-Merino, MD, PhD; Óscar Fernández, MD, PhD; Jens Kuhle, MD, PhD; Claudio Gobbi, MD; Amber Delmas, MD, PhD; Steve Simpson-Yap, PhD, MPH; Nupur Nag, PhD; Bassem Yamout, MD, FAAN; Nina Steinemann, PhD; Pierrette Seeldrayers, MD; Bénédicte Dubois, MD, PhD; Ingrid van der Mei, PhD; Alexander Stahmann, MSc; Jelena Drulovic, MD, PhD; Tatjana Pekmezovic, MD, PhD; Waldemar Brola, MD, PhD; Mar Tintore, MD, PhD; Nynke Kalkers, MD, PhD; Rumen Ivanov, MSc; Magd Zakaria, MD; Maged Abdel Naseer, MBBCh, MSc, MD, FAAN; Wim Van Hecke, PhD; Nikolaos Grigoriadis, MD, PhD; Marina Boziki, MD, PhD; Adriana Carra, MD; Mikolaj A. Pawlak, MD, PhD; Ruth Dobson, MRCP, PhD; Kerstin Hellwig, MD, PhD; Arlene Gallagher, PhD; Letizia Leocani, MD, PhD; Gloria Dalla Costa, PhD; Nise Alessandra de Carvalho Sousa, MD; Bart Van Wijmeersch, MD, PhD; Liesbet M. Peeters, PhD

Background: One of the major objectives of the Multiple Sclerosis Data Alliance (MSDA) is to enable better discovery of multiple sclerosis (MS) real-world data (RWD).

Methods: We implemented the MSDA Catalogue, which is available worldwide. The current version of the MSDA Catalogue collects descriptive information on governance, purpose, inclusion criteria, procedures for data quality control, and how and which data are collected, including the use of e-health technologies and data on collection of COVID-19 variables. The current cataloguing procedure is performed in several manual steps, securing an effective catalogue.

Results: Herein we summarize the status of the MSDA Catalogue as of January 6, 2021. To date, 38 data sources across five continents are included in the MSDA Catalogue. These data sources differ in purpose, maturity, and variables collected, but this landscaping effort shows that there is substantial alignment on some domains. The MSDA Catalogue shows that personal data and basic disease data are the most collected categories of variables, whereas data on fatigue measurements and cognition scales are the least collected in MS registries/cohorts.

Conclusions: The Web-based MSDA Catalogue provides strategic overview and allows authorized end users to browse metadata profiles of data cohorts and data sources. There are many existing and arising RWD sources in MS. Detailed cataloguing of MS RWD is a first and useful step toward reducing the time needed to discover MS RWD sets and promoting collaboration. *Int J MS Care.* 2021;23:261-268.

Real-world data (RWD), defined as data derived from a variety of origins related to outcomes in a heterogeneous patient population in a real-world setting, such as patient surveys, observational cohort studies, or registries, are increasingly used to address clinical questions related to multiple sclerosis (MS).[1] Because of the increasing awareness regarding the importance of using RWD, the number of MS RWD collection efforts is growing. Considerable RWD and biosamples are collected by registries or initiatives. The many existing and emerging MS RWD initiatives are distinct regarding inclusion criteria of people with MS, variables collected, frequency of data collection, organizational aspects, and documentation processes, among others.[2-4] However, the specific and valuable information describing, among others, the background, purpose, inclusion criteria, and variables collected in these existing RWD sources is usually not publicly available. Metadata contain information about the data

Note: Supplementary material for this article is available at ijmsc.org.

DOI: 10.7224/1537-2073.2021-006

and data services, such as content, quality, condition, and other characteristics of a data set. Developing a detailed metadata catalogue containing these descriptive data of existing and emerging MS RWD initiatives is, therefore, a useful first step to improve the findability and accessibility of RWD to improve awareness of existing and planned cohorts. In general, creating metadata catalogues contributes to open communication regarding data sources within scientific communities and thereby facilitates collaborative research and improves use of already existing RWD sources.[1,2]

Several RWD cataloguing efforts are listed in Table S1, which is published in the online version of this article at ijmsc.org. In 2010, the European Register for Multiple Sclerosis (EUReMS)[5,6] project, initiated and led by the European Multiple Sclerosis Platform (EMSP), was launched. It aimed to establish a network to compare longitudinally collected MS data in Europe. As part of the EUReMS project, the European Mapping Survey was performed in 2012.[3] To update and increase the knowledge of the MS registry data, the survey was rerun in 2017.[4] These mapping surveys provided an overview of data that are currently collected by participating MS registries, their governance, operational methods, and other factors. In 2018, a landscape analysis was published[2] revealing a significant number of largely uncoordinated parallel studies. This landscape analysis provides a useful high-level overview of 25 RWD sets across the globe and compares them based on some key features, including first participant enrollment date and geographic catchment of data.[2] More recently, the Secondary Progressive MS Research Collaboration Network, consisting of eight MS registries including more than 61,109 people with MS,[7] undertook a cataloguing effort to enable analysis of core variables collected in patients with secondary progressive MS in a real-world setting.[8] Outside Europe, the Multiple Sclerosis Metadata Collective (MSMDC), a collaborative effort of North American observational MS studies, was initiated by the US National Multiple Sclerosis Society (http://www.nationalmssociety.org) and the Consortium of Multiple Sclerosis Centers (CMSC; http://www.mscare.org). The MSMDC is using the Maelstrom Research cataloguing toolkit,[9] which already showcased their relevance in providing a valuable roadmap for conducting high-quality harmonization projects in the field of MS but also in other national and international initiatives. Indeed, a meta-analytic approach using the Maelstrom Research guidelines recently investigated employment status in people with MS across three MS registries.[17]

There are many other inspiring efforts outside the scope of MS showcasing the relevance of cataloguing (Table S1). The European Medical Information Framework (EMIF) was an Innovative Medicines Initiative (IMI) project that ran for 5.5 years (2013-2018) and successfully improved access to human health data via, among others, the provision of tools such as the EMIF Catalogue for consistent leveraging of available population-based and cohort-derived data sources to support novel research.[11] Another more

From the University MS Center, Hasselt-Pelt, Belgium (LG, TParciak, AP, BVW, LMP); Biomedical Research Institute (BIOMED) (LG, TParciak, AP, BVW, LMP) and Data Science Institute (LG, TParciak, AP, LMP), University of Hasselt, Diepenbeek, Belgium; University Medical Center Göttingen, Department of Medical Informatics, Germany (TParciak); ESAT-STADIUS, KU Leuven, Leuven, Belgium (AP); Accelerated Cure Project for MS, Waltham, MA, USA (RM, HS); Association of Multiple Sclerosis Societies of Croatia, Zagreb (TM); Center for Clinical Neuroscience, University Hospital Dresden, Germany (TZ); NeuroTransData Study Group, Neuburg/Donau, Germany (AB); Neurology Department, Hospital Universitario de CEMIC, Buenos Aires, Argentina (JIR); MS Center of Buenos Aires (CEMBA), Argentina (EC); Department of Neurology, Universidad Autonoma de Madrid, Spain (JAG-M); Neurology Service, Puerta de Hierro Hospital, Majadahonda, Madrid, Spain (JAG-M); University of Malaga, Department of Pharmacology, Spain (OF); Neurologic Clinic and Policlinic, MS Center and Research Center for Clinical Neuroimmunology and Neuroscience Basel (RC2NB), University Hospital Basel, University of Basel, Basel, Switzerland (JK); Multiple Sclerosis Center, Department of Neurology, Neurocenter of Southern Switzerland, Lugano, Switzerland (CG); Faculty of Biomedical Sciences, Università della Svizzera Italiana, Lugano, Switzerland (CG); Life Sciences Department, E*HealthLine.com, Inc (AD); Neuroepidemiology Unit, Melbourne School of Population and Global Health, The University of Melbourne, Australia (SS-Y, NN); Multiple Sclerosis Center, American University of Beirut Medical Center, Lebanon (BY); Data Center of the Swiss Multiple Sclerosis Registry, Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Switzerland (NS); Department of Neurology, CHU de Charleroi, Charleroi, Belgium (PS); Department of Neurosciences, Laboratory for Neuroimmunology, KU Leuven, Leuven, Belgium (BD); Leuven Brain Institute KU Leuven, Leuven, Belgium (BD); Department of Neurology, University Hospitals Leuven, Leuven, Belgium (BD); Menzies Institute for Medical Research, University of Tasmania, Hobart TAS, Australia (IvdM); German MS-Registry, MS Forschungs- und Projektentwicklungs-gGmbH, Hannover, Germany (AS); Clinic of Neurology, Clinical Center of Serbia, Belgrade, Serbia (JD); Institute of Epidemiology, Faculty of Medicine, University of Belgrade, Belgrade, Serbia (TPekmezovic); Collegium Medicum, Jan Kochanowski University, Kielce, Poland (WB); Centre d'Esclerosi Múltiple de Catalunya (Cemcat), Edifici Cemcat, Hospital Universitari Vall d´Hebron, Barcelona, Spain (MT); Department of Neurology, OLVG, and Department of Neurology, Amsterdam UMC, Location VUMC, Amsterdam, the Netherlands (NK); PMA - Pharma Marketing Advisors, Ltd, Sofia, Bulgaria (RI); Department of Neurology, Ain Shams University, Egypt (MZ); Department of Neurology, Cairo University, Egypt (MAN); icometrix, Leuven, Belgium (WVH); Second Neurological University Department, Multiple Sclerosis Center, Aristotle University of Thessaloniki, AHEPA General University Hospital, Thessaloniki Greece (NG, MB); MS Center Hospital Britanico, Buenos Aires, Argentina (AC); Department of Neurology and Cerebrovascular Disorders, Poznan University of Medical Sciences, Poznan, Poland (MAP); Wolfson Institute of Preventive Medicine, Charterhouse Square, London, UK (RD); Department of Neurology, Katholisches Klinikum, St Josef Hospital, Ruhr University Bochum, Bochum Germany (KH); Clinical Practice Research Datalink (CPRD), Medicines and Healthcare Products Regulatory Agency (MHRA), London, UK (AG); Clinical Neurology Unit, San Raffaele University, Milan, Italy (LL, GDC); Hospital Universitário Getulio Vargas, Manaus/Amazonas, Brazil (NAdCS); REDONE – Brazilian Registry of Multiple Sclerosis and Neuromyelitis Optica Spectrum Disorders, São Paulo, Brazil (NAdCS); and Noorderhart, Rehabilitation and MS Center, Pelt, Belgium (BVW). *Correspondence:* Liesbet Peeters, PhD, Agoralaan Building C, University of Hasselt, 3590 Diepenbeek, Belgium; e-mail: liesbet.peeters@uhasselt.be.

recent and still ongoing IMI project (2019-2024) is ConcePTION.[12] This project is focused on data collection from pregnant and breastfeeding women who are otherwise mostly excluded from research studies because of safety concerns. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) Resource Database,[13] launched in 2008, aims to facilitate researcher access to electronic databases in Europe by inviting data custodians to provide descriptions of their core data.[14] As part of the Open Government Initiative, the US Federal Government published important data sets to the public using an online portal, HealthData.gov, to increase collaboration, transparency, consumer participation, and research.[16]

The Multiple Sclerosis Data Alliance (MSDA)[18] is a global multistakeholder collaboration working to accelerate insights for innovative care and treatment for people with MS. The MSDA aims to raise awareness of MS RWD, build a multistakeholder MS data community, promote trustworthiness about the use of RWD, and tackle technical challenges related to RWD. The MSDA Catalogue provides descriptive metadata and, if reported by the custodians, number of patients, visits, and collected data of contributing MS data sources. The Catalogue enables dynamic landscaping of the existing data within different registries and initiatives spread throughout the world, thereby allowing researchers to determine which data sources offer relevant or high-impact variables in a specific context. Because MS is a quite rare, very heterogenous, and, in many ways, poorly understood disease that should be treated in a multidisciplinary manner and for which multiple disease-modifying treatments are available, it is important to answer certain research questions by using RWD. To generate trustworthy results, a data set of sufficient size and data quality is required, implying that collaboration between different MS RWD sources is often of essence.[1,9] The MSDA Catalogue, which is available worldwide, aims to reduce the time needed for researchers to discover MS RWD sets and, therefore, to facilitate collaborative research.

Methods

With the MSDA Catalogue, we worked on top of the efforts performed in two other large initiatives: the EMIF and the updated European mapping exercise.[4] The EMIF Catalogue[10] is built around disease communities. The MSDA used to be one of these disease communities within the EMIF platform, but recently the MSDA Catalogue got its own personal platform (https://msda.emif-catalogue.eu/), which can be browsed for relevant MS data sets by interested parties after creating an account. The questionnaire of the European mapping survey was digitized to serve as a basis for the MSDA Catalogue, collecting for each data source its organizational information, background/purpose of the registry, inclusion criteria of patients and centers, information regarding documentation processes, data collected, quality control mechanisms, governance, and status of the registry (October 2020). In November 2020, the questionnaire was extended with high-level questions regarding e-health technologies and COVID-19. More detailed information about the domains of the questionnaire can be found in Table S2. The platform is easily adaptable, which means that the list of questions can be updated as needed.

Registry recruitment for the MSDA Catalogue started in June 2019 and has been ongoing. Besides custodians of national registries, researchers who have established unique MS research cohorts were approached to contribute to the Catalogue. The list of candidate registries and data sources was created through a combination of literature search, conference attendance, through the MSDA network, or via their collaborating partners. Since June 2019, 62 data custodians have been contacted via e-mail or in person and asked to complete the e-questionnaire. All data custodians in the MSDA community can edit and update the metadata of their data sources whenever needed. We strive to update the metadata regularly, with an update at least once a year. In addition, the e-questionnaire is regularly evaluated and adapted based on feedback.

Results

Currently, 38 registries are enrolled in the MSDA Catalogue (Table 1). These initiatives collect data from people with MS across five continents and differ in their origin, inclusion criteria, data collected (clinical outcomes, patient-reported outcomes, biospecimens, imaging), and duration of follow-up. There are 15 languages used among the 38 registries, but English is the most common, used in 21 of them. Because global initiatives, as well as separate entities that share their data with these international registries, are part of the Catalogue, there is a certain amount of overlap of people with MS. However, we considered it useful to publish all these initiatives because of the differences between the data sets and organizations.

The participating registries were initiated between 1956 and 2020, and sample sizes range from a few hundred to several thousand (>75,000) people with MS. Of the 38 registries, 18 (47%) are kept by academic/research institutions, eight (21%) by health care organizations, six (16%) by patient organizations (although patient organizations are involved in 21 registries [55%]), and 11 (29%) by other institutions or organizations, such as the government or private companies. In 34 of the 38 registries (89%), health care services are the sources of data collection, and more specifically, general neurologists (n = 18; 47%), child neurologists/pediatricians (n = 8; 21%), rehabilitation units (n = 8; 21%), and general care (n = 4; 11%). Public authorities/administrative databases and MS societies/patient organizations are reported eight (21%) and six (16%) times, respectively, as sources of data collection.

Participants were included for data collection if MS was diagnosed based on the McDonald criteria in 30 of the 38 registries (79%), of which 15 (39%) additionally included people with MS according to the Poser criteria. Twenty-two registries (58%) include participants with possible MS/clinically isolated syndrome, and eight registries (21%) collect data on patients with self-reported MS that is not clinically validated.

Table 1. Registries that are currently part of the MSDA Catalogue

Name-sharing initiative (abbreviation)	Geographic coverage	Language	Year database initiated	Current approximate no. of registered MS patients
Accelerated Cure Project Repository (ACP)	United States	English	2005	2500
Association of Multiple Sclerosis Societies of Croatia (AMSSC)	Croatia	Croatian	2006	3130
Australian MS Longitudinal Study (AMSLS)	Australia	English	2002	3000
Belgian Treatments in MS (BELTRIMS)	Belgium	English	2012	2028
Centre d'Esclerosi Múltiple de Catalunya (Cemcat) cohort	Catalonia	Spanish (working on switching variables to English)	1995	3964
Clinical Practice Research Datalink (CPRD)	United Kingdom	English	1987	16,000,000[a]
E*HealthLine.com	Europe, United States, Asia	English, other languages	2014	20,000
German Multiple Sclerosis and Pregnancy Registry (DMSKW)	Germany	German	2006	2500
Health Outcomes and Lifestyle In a Sample of people with Multiple sclerosis Study (HOLISM)	66 countries	English	2012	3039
Hellenic Academy of Neuroimmunology (HELANI) – COVID-19 questionnaire for patients with multiple sclerosis and their carers	Greece	Greek, English	2020	490
icompanion	Global	English, French, German, Dutch, Spanish, Italian	2020	3500
iConquerMS - The People-Powered Research Network	United States	English	2014	7000
Italian MS Register	Italy	Italian, English (in preparation)	2000	68,106
Middle East North Africa Committee for Treatment and Research in MS (MENACTRIMS)	Middle East, North Africa	English	2016	10,000
MS Data Connect (MSDC)	Belgium	Dutch, English	2017	900
MS Network - Egypt Registry	Egypt	English	2015	5245
MS Society of Serbia	Serbia	Serbian	1996	2725
MSBase	Global	English, German, Italian, Spanish, French	2004	75,600
MSDS Study Group	Germany	German, English	1998	10,000
MSGene	Poland	Polish, English	2013	217
MS-Register der DMSG, Bundesverband e. V. (MS-Register by the German MS-Society)	Germany	German	2014 (2001)	32,643 (since 2014)
MSRegistratie	The Netherlands	Dutch	2017	950
NeuroTransData Registry Database (NTD)	Germany	German	2008	25,001
Norwegian MS Registry & Biobank	Norway	Norwegian	1998	Unknown
Optimise MS Database	United Kingdom	English	2019	1800
Polish MS Registry (RejSM)	Poland	Polish	2011	10,500
RADAR-CNS	Europe, United States	English	2020	503
REDONE.br	Brazil	Portuguese, English	2018	1500
Registro Argentino de Esclerosis Múltiple (RelevarEM)	Argentina	Spanish	2016	3256
Registro Español de Esclerosis Múltiple	Spain	Spanish	2020	Ongoing
ReMuS	Czech Republic	Czech	2013	16,988
SmartMS	Bulgaria	Bulgarian	2016	2231
Survey on the impact of COVID-19 in patients with multiple sclerosis in Latin America	Latin America	Spanish	2020	774
Swedish Neuro Registries - MS	Sweden	Swedish	1997	20,836
Swiss Multiple Sclerosis Cohort Study	Switzerland	English	2012	1460
Swiss Multiple Sclerosis Registry	Switzerland	German, French, Italian	2016	2400
The Danish Multiple Sclerosis Registry	Denmark	Danish, English	1956	29,000
United Kingdom MS Register	United Kingdom	English	2011	16,000

[a]Total number of patients, not exclusively people with MS.

In the MSDA Catalogue, data source representatives indicate who performs the primary documentation to include data in their database. Multiple answers are possible to select, and the observational analysis showed that in 31 of the participating registries (82%), primary documentation is entered by a neurologist and in 18 (47%) (additionally) by a medical assistant or nurse. Patients perform the primary documentation in 16 of 38 registries (42%), whereas other physicians (eg, paramedical staff) perform the documentation process in two registries (5%). Nine registries (24%) reported other sources as being responsible for primary documentation, which was further specified and included a PhD pharmacist, a data manager, a study coordinator, a neurology resident, and a research assistant.

In 12 of 38 registries (32%), data are collected on paper forms, but only in one of these are paper forms the only medium of data collection. Local data acquisition systems (manual merging of data sets from different sources) were reported by ten registries (26%) and remote data entry systems (eg, Web-based data capture, central storage of data sets from different sources) were reported by 31 registries (82%) as their data collection mode.

Data entry is triggered by fixed consultation dates, mostly yearly or twice a year, in 17 registries (45%) and by patients visiting a data-supplying center in ten (26%). Events within the context of MS disease trigger data entry in 14 registries (37%), of which six (16%) reported events within the context of health care as triggering factors as well.

In 33 registries (87%), different data sets of one patient, acquired in one or multiple centers at different time points, can be linked together, supporting longitudinal surveillance of the patients.

"Data that are collected" is one of the sections in the questionnaire that feeds the MSDA Catalogue. Within this part, data custodians can declare which data they collect in their registry, without sharing patient-level data. The "data that are collected" section is subdivided into 17 categories: personal data, basic disease data, relapses, disability, cognition scales, treatments (relapse treatment, disease-modifying therapies [DMTs], symptomatic treatments), magnetic resonance imaging, paraclinical measures, patient-derived measures, depression, fatigue, comorbidities, socioeconomic data, societal services, health care utilization, e-health technologies, and COVID-19 and MS. For each category, numerous variables (or tests generating variables) are listed for which the data custodians can indicate whether these variables are collected in their initiative. It is also possible for registries to add additional categories and variables in "other categories."

Detailed information regarding the data collected can be found in Table S3. The categories with the highest coverage regarding data collection are personal data (containing date of birth and sex) (99%), basic disease data (eg, time of disease onset, time of diagnosis, disease course) (83%), and relapses (eg, date of relapse, corticosteroid treatment of relapse, relapse symptoms) (71%).

In terms of variables within the categories, all 38 registries collect date or year of birth, whereas sex and time of disease onset are collected by 37 registries (97%). Time of diagnosis and current DMTs are collected by 36 registries (95%), and disease course is collected by 34 (89%). Expanded Disability Status Scale (EDSS) score and start date of current therapy are each collected by 33 registries (87%) and past DMTs by 32 registries (84%).

Fatigue (eg, the Krupp Fatigue Severity Scale, Modified Fatigue Impact Scale, and Fatigue Impact Scale) (12%), cognition scales (eg, the Symbol Digit Modalities Test, Paced Auditory Serial Addition Test, and Brief International Cognitive Assessment for MS) (14%), and patient-derived measures (eg, the EQ-5D and MS Impact Scale) (16%) are the categories least covered by the registries and initiatives that have shared their metadata with the MSDA Catalogue.

However, within each specific category there is often great variability in coverage between certain variables that belong to that category. For example, in the disability category, EDSS score is collected by 87% of the registries (n = 33), whereas the Multiple Sclerosis Functional Composite is collected by only 42% (n = 16). Likewise, in the treatment category, current DMT is collected by 95% of the registries (n = 36), but treatment satisfaction–patient-reported is collected by 21% (n = 8). Therefore, the average coverage of the disability and treatments categories is 53% and 62%, respectively, despite the high coverage of EDSS score and DMT variables.

Data regarding COVID-19 and MS are collected by 26 of the registries (68%) in the MSDA Catalogue. These 26 registries specified the source of COVID-19 in MS data collection, in which a distinction between patient-reported (n = 8) and clinician-reported (n = 8) is made. Ten registries collect both patient- and clinician-reported COVID-19 in MS data. Furthermore, of the 26 initiatives collecting data on COVID-19 and MS, 15 registries indicated that they participated as a data partner in the global data sharing initiative of the MSDA in collaboration with the MS International Federation.[18]

Currently, 11 of the registries (29%) collect data using e-health technologies, which could be further specified as social media, mobile apps, wearable devices, and technical tools for neurorehabilitation. Of these 11 registries, eight and five indicated using mobile apps and wearable devices, respectively. Social media (n = 2) and technical tools for neurorehabilitation (n = 1) are currently less commonly used.

Of the participating registries, nine (24%) perform quality control manually, 11 (29%) automatically, 14 (37%) using both manual and automatic methods, and two (5%) not at all. Two registries (5%) did not specify whether they use manual or automatic quality control mechanisms but indicated that quality control was implemented.

Thirty-six participating registries (95%) report approval by local authorities, further specified as data protection authorities (n = 17; 45%), ethics committee (n = 32; 84%), or other (n = 5; 13%). Two registries (5%) report no approval by local authorities. Written informed consent is obtained by 29 registries (76%), whereas seven (18%) report other forms of informed consent (eg, digital informed consent).

Discussion

Many other organizations or initiatives (Table S4) are or have been focusing on developing guidelines or recommendations to optimize collaborative research using different RWD sets. The MSDA Catalogue of existing MS cohorts and registries was developed to speed up the discovery of appropriate MS RWD sources. Sources of MS RWD differ in size, purpose, maturity, and collection of variables. This has already been shown in previous landscaping exercises[2-4] and is again confirmed in this initiative. The differences among existing MS RWD sources highlights the need to align data collection efforts as much and as early as possible. Bringing together data from multiple sources for collaboration is facilitated by agreeing on a core data set dictionary (containing the variables that should minimally be collected, their format, and data type) for either prospective data collection or retrospective data harmonization. The importance of agreeing on a core data set has been clearly illustrated previously. Indeed, in the COVID-19 in MS global data sharing initiative,[31] a COVID-19 in MS core data set has been defined by a global taskforce. Subsequently, multiple MS registries and cohorts implemented this core data set to start data collection on COVID-19 in MS or used the core data set dictionary as a target data set to transform their data to. Sharing the collected data resulted in a data set of more than 10,000 records, which were used to generate reliable answers to urgent questions during the COVID-19 pandemic.[32] Another example showcasing the relevance of agreeing on a core data set is the protocol for postauthorization safety studies developed within the framework of the efforts performed by the Big MS Data Network.[22] There are several existing common data models (CDMs) that are not MS specific. A CDM defines unambiguously the semantic and syntactic representation of data, implying that a shared language for the data is provided. Examples of existing CDMs include the Observation Medical Outcome Partnership (OMOP) CDM,[33] the National Patient-Centered Clinical Research Network (PCORnet) CDM,[34] and the Sentinel CDM (SCDM).[35] The OMOP CDM is also used within the European Health Data & Evidence Network (EHDEN; https://www.ehden.eu/), an IMI initiative that started in 2018 and will operate until 2024 and commits to developing a federated network of data sources that use a standardized CDM to improve research collaborations and health decisions and care.

Cataloguing efforts can contribute to increasing the understanding about the different data source structures and formats that are used. However, creating data dictionaries and regularly sharing metadata is a time-consuming task for data custodians, specifically when registries are involved in several parallel cataloguing efforts, requiring repeated manual entry of identical or similar metadata. As described previously herein and summarized in Table S1, there are already many emerging and existing cataloguing efforts. Alignment between different cataloguing pipelines will allow data custodians to interact with multiple initiatives if they want to do so while reducing the need for repeated manual entry. For example, the MSDA is currently looking into how to connect the MSDA cataloguing pipeline to the Maelstrom Research pipeline.

There are already so-called automated data screening tools available to speed up the process of generating metadata about a data source, such as the data format and type. For example, WhiteRabbit[36] is a software tool developed within the framework of the Observational Health Data Sciences and Informatics (OHDSI) program. The OHDSI program is a multistakeholder, interdisciplinary partnership that seeks to uncover the value of observational health data through large-scale analytics. The main function of WhiteRabbit is to perform a scan of a data source, providing detailed information about the tables, fields, and values that appear in a field. Another potentially interesting tool/software package with a similar output as WhiteRabbit is Pandas Profiling.[37] Pandas Profiling library can assist in exploring, cleaning, and processing of data stored as a data

PRACTICE POINTS

- The Multiple Sclerosis Data Alliance (MSDA) is a global multistakeholder collaboration that aims to raise awareness about the importance of research using MS real-world data (RWD) and seeks to enable better discovery and access to MS RWD.
- The MSDA launched the MSDA Catalogue in 2019. It allows end users with particular study requirements or research questions to browse metadata profiles of MS RWD cohorts.
- The authors invite all MS clinicians and researchers to actively contribute to MS RWD collection efforts and to share metadata on their data collections.

frame. Using this library, profile reports can be created in different formats, such as HTML or JSON. These reports include descriptive statistics, quantile statistics, the possible correlation between the objects in the data frame, and some metadata information regarding values stored in the data frame. Automatic tools and methods such as WhiteRabbit and Pandas Profiling are particularly useful to generate descriptive information regarding patient-level data sources and to create data dictionaries, which is crucial to start collaborative research.

The goal of the MSDA is, indeed, more sophisticated than the Catalogue in its current format. We will investigate further how we can evolve from a manually operated catalogue to a (semi-)automated tool that can be deployed to solve research questions faster and generate real-world evidence to advance care and treatment of people with MS.

In conclusion, the Web-based MSDA Catalogue provides strategic overview and allows authorized end users to browse metadata profiles of data cohorts and data sources. There are many existing and arising RWD sources in MS. They differ in purpose, maturity, and the variables they collect. The individuals who administer most data registries are in favor of data sharing and are willing to work diligently to achieve harmonized data sets. However, data registries are long-term collaborative scientific efforts developed according to their own guidelines, laws, and governance of data access, which affects open access and sharing of the collected data directly. In addition, cataloguing and data harmonization are time-consuming, labor-intensive, and, therefore, expensive activities, whereas funding sources are often scarce. However, this landscaping effort shows that there is substantial alignment on some domains. Cataloguing can be useful in discovering suitable RWD sources to solve specific research questions and is a potential first step toward speeding up collaborative research projects. Indeed, deep knowledge and understanding about the data source structure and format are needed to facilitate and speed up harmonization efforts that could support collaborative research initiatives. We invite all MS clinicians and researchers to actively contribute to MS RWD collection efforts and to regularly share data dictionaries and metadata on their data cohorts. □

Acknowledgments: We thank the sponsors of the MSDA and all the data custodians who completed and shared the questionnaire with the MSDA community. Special thanks to all the people involved in the data collection and management of the different data sources and registries.

Financial Disclosures: Dr Geys is funded by the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie Vlaanderen." Ms Parciak and Mr Pirmani are funded by the Fonds voor Wetenschappelijk Onderzoek (FWO) (project R4859). Dr McBurney is the chief research officer and Ms Schmidt is the vice president of scientific operations of the Accelerated Cure Project for MS (ACP). Dr McBurney has received consulting payments from EMD Serono, which have been donated to the ACP. The ACP has received grants, collaboration funding, payments for use of assets, or in-kind contributions from the following companies: EMD Serono, Novartis, Sanofi/Genzyme, Biogen, Genentech, AbbVie, Octave, GlycoMinds, Pfizer, MedDay, AstraZeneca, Teva, Mallinckrodt, MSDx, Regeneron Genetics Center, BC Platforms, and Bristol Myers Squibb (Celgene). The ACP has also received funding from the Patient-Centered Outcomes Research Institute, the National MS Society, and the Italian MS Society. Dr Rojas has received reimbursement for developing educational presentations, educational and research grants, consultation fees, and travel stipends from Biogen, Genzyme, Merck, and Novartis. Dr García-Merino has received honoraria for travel expenses, lecturing, or consulting from Merck, Biogen, Bayer, Teva, Roche, Novartis, Genzyme, and Almirall and research grants from Biogen and Novartis. Dr Fernández has received travel support and/or research grants and/or lecture fees and/or advisory services from Merck, Bayer, Biogen, Teva, Sanofi Genzyme, Novartis, Roche, Almirall, Allergan, Orizon, and Araclon. Dr Gobbi's employer (Department of Neurology, Regional Hospital Lugano [EOC], Switzerland) receives financial support for speaking or educational, research, or travel grants from AbbVie, Almirall, Biogen, Celgene, Sanofi, Merck, Novartis, Teva, and Roche. The Neuroepidemiology Unit and HOLISM-related work (Drs Simpson-Yap and Nag) was supported by anonymous philanthropic funders and Mr Wal Pisciotta. Dr Yamout has received speaker honoraria from Bayer, Biogen, Genesis Pharma, Genpharm, Genzyme, Merck, and Novartis; research grants from Bayer, Biogen, Merck, Novartis, and Pfizer; and advisory board honoraria from Bayer, Biogen, Genzyme, Merck, Novartis, and Genpharm. The Swiss MS Registry (Dr Steinemann) is funded by a grant from the Swiss Multiple Sclerosis Society. Dr Seeldrayers has been responsible for BELTRIMS in the frame of the Belgian Study Group for Multiple Sclerosis; in the frame of her clinical activities, she has received travel grants and honoraria for lectures and participation in advisory boards from Biogen, Bayer, Merck, Teva, Sanofi Genzyme, Novartis, and Roche. Dr Dubois is a clinical investigator of the Research Fund Flanders (FWO-Vlaanderen). Mr Stahmann is the lead of the German MS Registry, which receives funding from a range of public and corporate sponsors, recently including the German Innovation Fund, German MS Trust, German MS Society, Biogen, Celgene (Bristol Myers Squibb), Merck, Novartis, Roche, and Sanofi. Dr Drulovic has received travel support and/or research grants and/or lecture fees and/or advisory services from Novartis, Bayer, Merck, Sanofi Genzyme, Roche, Teva, Medis, Hemofarm, and Medtronic. Dr Pekmezovic has received travel support and/or research grants and/or lecture fees and/or advisory services from Novartis, Merck, Sanofi Genzyme, Roche, Teva, and Hemofarm. Dr Tintore discloses consulting services and speaking honoraria from Almirall, Bayer Schering Pharma, Biogen, Genzyme, Janssen, Merck-Serono, Novartis, Roche, Sanofi-Aventis, Viela Bio, and Teva and research grants and support from Carlos III Health Institute, Foundation Genzyme, Foundation *Salud 2000*, Biogen, and Novartis. Dr Kalkers received speaking fees and consulting fees from Novartis, Teva, and Sanofi Genzyme. The Dutch MS Registry is funded by a project from ZonMW and co-financed by Biogen, Celgene, Merck, Mylan, Novartis, Roche, Sanofi Genzyme, and Teva. Dr Zakaria has received fees for lectures and advisory board meetings from Merck, Roche, Sanofi, Biogen, Novartis, and Bayer. Dr Naseer received speaker honoraria and served on scientific advisory boards for Novartis, Roche, Merck, Sanofi Genzyme, and Bayer. Dr Grigoriadis has received travel support and/or honoraria and/or research grants and/or lecture fees and/or advisory services and/or consulting fees from Novartis, Bayer, Merck, Genesis, Sanofi, Specifar, Roche, Biogen, Teva, Mylan, Merck Serono, Sanofi Genzyme, Celgene, and ELPEN. Dr Boziki has received travel support and/or research grants and/or lecture fees and/or advisory services from the Hellenic Foundation for Research and Innovation, the Ministry of Education's Education and Lifelong Learning Program, the Hellenic Neurological Society, the Hellenic Academy of Neuroimmunology, Biogen, Novartis, Teva, Bayer, Genesis Pharma, Sanofi, Specifar, Roche, and Merck. Dr Pawlak has received grants from the Polish National Science Centre during the conduct of the study as well as personal fees from Roche, Novartis, Merck, and Biogen outside the submitted work. Dr Dobson has received travel support and/or research grants and/or consulting fees from Biogen, Roche, Celgene, Sanofi Genzyme, Teva, and Merck. Dr Hellwig has received travel support and/

or research grants and/or lecture fees and/or advisory services from Novartis, Bayer, Merck, Sanofi Genzyme, Roche, Biogen, Teva, and Mylan. Dr Leocani has received consulting fees and/or research support from Biogen, Merck KGgA, Novartis, Hoffman-La Roche, Almirall SpA, and Bristol Myers Squibb. Dr Van Wijmeersch has received travel support and/or research grants and/or lecture fees and/or advisory services from Almirall, Biogen, Bristol Myers Squibb, Janssen Pharmaceutica, Merck, Novartis, Roche, Sanofi, and Teva. Dr Peeters is chair of the MSDA, which receives income from a range of corporate sponsors, including Biogen, Bristol Myers Squibb, Janssen, Merck KGaA, Mylan, Novartis, QMENTA, and Roche and has received funding from the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie Vlaanderen." The other authors declare no conflicts of interest.

Funding/Support: We thank our sponsors. The MSDA receives income from a range of corporate sponsors, recently including Biogen, Bristol Myers Squibb, Janssen Pharmaceuticals, Merck, Mylan, Novartis, QMENTA, and Roche.

References

1. Cohen JA, Trojano M, Mowry EM, Uitdehaag BM, Reingold SC, Marrie RA. Leveraging real-world data to investigate multiple sclerosis disease behavior, prognosis, and treatment. *Mult Scler*. 2020;26:23-37.
2. Bebo BF, Fox RJ, Lee K, Utz U, Thompson AJ. Landscape of MS patient cohorts and registries: recommendations for maximizing impact. *Mult Scler*. 2018;24:579-586.
3. Flachenecker P, Buckow K, Pugliatti M, et al. Multiple sclerosis registries in Europe: results of a systematic survey. *Mult Scler*. 2014;20:1523-1532.
4. Glaser A, Stahmann A, Meissner T, et al. Multiple sclerosis registries in Europe: an updated mapping survey. *Mult Scler Relat Disord*. 2019;27:171-178.
5. European Register for Multiple Sclerosis (EUReMS) report: A collaborative initiative to improve MS research and policy across Europe. EMSP. Accessed: October 2021. https://emsp.org/resources/eurems-report/
6. The Lancet Neurology. Making EUReMS count for people with multiple sclerosis. Editorial. *Lancet Neurol*. 2011;10:865.
7. Glaser A, Forsberg L, Manouchehrinia A, et al. Objective classification methods result in an increased proportion of secondary progressive multiple sclerosis in five patient registries. Poster presented at: MSVirtual2020; September 11-13, 2020. https://www.msregister.de/fileadmin/resources/public/documents/publications/poster/MSVirtual2020_P0482_Anna_Glaser.pdf
8. Glaser A, Trojano M, Laffaldano P, et al. Opportunities and challenges for conducting research on secondary progressive multiple sclerosis across international multiple sclerosis registries through a research network collaboration. Poster presented at: ECTRIMS 2019 Congress; September 11-13, 2019; Stockholm, Sweden. https://onlinelibrary.ectrims-congress.eu/ectrims/2019/stockholm/278744/jan.hillert.opportunities.and.challenges.for.conducting.research.on.secondary.html
9. Bergeron J, Doiron D, Marcon Y, Ferretti V, Fortier I. Fostering population-based cohort data discovery: the Maelstrom Research cataloguing toolkit. *PLoS One*. 2018;13:e0200926.
10. EMIF Catalogue. Accessed November 27, 2020. https://emif-catalogue.eu
11. Lovestone S, Consortium E. The European medical information framework: a novel ecosystem for sharing healthcare data across Europe. *Learn Health Syst*. 2020;4:e10214.
12. ConcePTION. Innovative Medicines Initiative. Accessed January 7, 2021. https://www.imi.europa.eu/projects-results/project-factsheets/conception
13. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). European Medicines Agency. Accessed December 4, 2020. https://www.ema.europa.eu/en/partners-networks/networks/european-network-centres-pharmacoepidemiology-pharmacovigilance-encepp
14. Kurz X, Perez-Gutthann S, ENCePP Steering Group. Strengthening standards, transparency, and collaboration to support medicine evaluation: ten years of the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). *Pharmacoepidemiol Drug Saf*. 2018;27:245-252.
15. Multiple Sclerosis Metadata Collective: a collaborative effort of North American observational studies in multiple sclerosis. Maelstrom Research. Accessed October 2021. https://www.maelstrom-research.org/network/msmdc
16. Marc DT, Beattie J, Herasevich V, Gatewood L, Zhang R. Assessing metadata quality of a federally sponsored health data repository. *AMIA Annu Symp Proc*. 2016;2016:864-873.
17. Salter A, Stahmann A, Ellenberger D, et al. Data harmonization for collaborative research among MS registries: a case study in employment. *Mult Scler*. 2021;27:281-289
18. Peeters LM, Parciak T, Kalra D, et al. Multiple Sclerosis Data Alliance: a global multi-stakeholder collaboration to scale-up real world data research. *Mult Scler Relat Disord*. 2021;47:102634.
19. Guideline on registry-based studies - launch of public consultation. European Medicines Agency. Accessed December 3, 2020. https://www.ema.europa.eu/en/news/guideline-registry-based-studies-launch-public-consultation
20. Patient Registries. 2015. European Medicines Agency. Accessed December 3, 2020. https://www.ema.europa.eu/en/human-regulatory/post-authorisation/patient-registries#-use-of-patient-disease-registries-for-regulatory-purposes-section
21. Multiple sclerosis workshop - registries initiative. European Medicines Agency. Published July 7, 2017. Accessed December 3, 2020. https://www.ema.europa.eu/en/events/multiple-sclerosis-workshop-registries-initiative
22. Hillert J, Trojano M, Magyari M, et al. Big Multiple Sclerosis Data: a registry basis for post authorization safety studies (PASS) for multiple sclerosis. Poster presented at: ECTRIMS 2019 Congress; September 11-13, 2019; Stockholm, Sweden. https://onlinelibrary.ectrims-congress.eu/ectrims/2019/stockholm/278582/jan.hillert.big.multiple.sclerosis.data.-.a.registry.basis.for.post.html
23. HMA-EMA Joint Big Data Taskforce Phase II report: 'Evolving Data-Driven Regulation.' European Medicines Agency. Accessed December 3, 2020. https://www.ema.europa.eu/en/documents/other/hma-ema-joint-big-data-taskforce-phase-ii-report-evolving-data-driven-regulation_en.pdf
24. The ENCePP Code of Conduct. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. Accessed December 3, 2020. http://www.encepp.eu/code_of_conduct/documents/ENCePPCodeofConduct.pdf
25. Zaletel M, Kralj M, eds. Methodological guidelines and recommendations for efficient and rational governance of patient registries. Accessed December 3, 2020. https://ec.europa.eu/health/sites/health/files/ehealth/docs/patient_registries_guidelines_en.pdf
26. The Lancet Neurology. Patient-reported outcomes in the spotlight. Editorial. *Lancet Neurol*. 2019;18:981. doi:10.1016/S1474-4422(19)30357-6
27. European Health Data Space. European Commission. Accessed December 3, 2020. https://ec.europa.eu/health/ehealth/dataspace_en
28. LaRocca NG, Hudson LD, Rudick R, et al. The MSOAC approach to developing performance outcomes to measure and monitor multiple sclerosis disability. *Mult Scler*. 2018;24:1469-1484.
29. Sastre-Garriga J, Pareto D, Battaglini M, et al. MAGNIMS consensus recommendations on the use of brain and spinal cord atrophy measures in clinical practice. *Nat Rev Neurol*. 2020;16:171-182.
30. Chalon PX, Kraemer P. EUnetHTA information management system: development and lessons learned. *Int J Technol Assess Health Care*. 2014;30:514-520.
31. Peeters LM, Parciak T, Walton C, et al. COVID-19 in people with multiple sclerosis: a global data sharing initiative. *Mult Scler*. 2020;26:1157-1162.
32. Simpson-Yap S, De Brouwer E, Kalincik T, et al: Associations of disease-modifying therapies with COVID-19 severity in multiple sclerosis. *Neurology*. October 5, 2021. Online ahead of print. doi: 10.1212/WNL.0000000000012753
33. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc*. 2012;19:54-60.
34. Klann JG, Abend A, Raghavan VA, Mandl KD, Murphy SN. Data interchange using i2b2. *J Am Med Inform Assoc*. 2016;23:909-915.
35. Sentinel. Accessed December 3, 2020. https://www.sentinelinitiative.org/
36. WhiteRabbit for ETL design. Observational Health Data Sciences and Informatics. Accessed January 8, 2021. https://www.ohdsi.org/analytic-tools/whiterabbit-for-etl-design/
37. Pandas-profiling. GitHub. Accessed January 8, 2021. https://github.com/pandas-profiling/pandas-profiling