ANNALS OF GEOPHYSICS, 65, 3, DM318, 2022; doi:10.4401/ag-8790
OPEN ACCESS

The EPOS Multi-Scale Laboratories: A FAIR Framework for Stimulating Open Science Practice across European Earth Sciences Laboratories

Kirsten Elger[*,1], Geertje ter Maat[2], Rita Caldeira[3], Corrado Cimarelli[4], Fabio Corbi[5], Stephane Dominguez[6], Martyn Drury[2], Francesca Funiciello[7], Otto Lange[8], Audrey Ougier-Simonin[9], Matthias Rosenau[1], Richard Wessels[2], Ernst Willingshofer[2], Aldo Winkler[10]

(1) GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany
(2) Utrecht University, Princetonlaan 8a, 3584CB Utrecht, The Netherlands
(3) Laboratorio Nacional de Energia e Geologia, Estrada da Portela Bairro do Zambujal Apartado 7586, Alfragide 2610-999 Amadora Portugal
(4) Ludwig-Maximilians-Universität München, Theresienstrasse 41, 80333 Munich, Germany
(5) Istituto di Geologia Ambientale e Geoingegneria – CNR c/o Dipartimento di Scienze della Terra, Sapienza Università di Roma, Rome, Italy
(6) Géosciences Montpellier, Université de Montpellier, Place Eugène Bataillon, 34095 Montpellier Cedex 05, France
(7) Dipartimento Scienze, Università Roma TRE, Largo San Leonardo Murialdo 1, 00146 Rome, Italy
(8) Utrecht University Library, Heidelberglaan 3, 3584 CS, The Netherlands
(9) British Geological Survey, Nicker Hill, Keyworth, United Kingdom
(10) Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy
[+] the MSL community

Article history: received January 4, 2022; accepted April 27,2022

Abstract

The Multi-scale Laboratories (MSL) are a network of European laboratories bringing together the scientific fields of analogue modeling, paleomagnetism, experimental rock and melt physics, geochemistry and microscopy. MSL is one of nine (see below) Thematic Core Services (TCS) of the European Plate Observing System (EPOS). The overarching goal of EPOS is to establish a comprehensive multidisciplinary research platform for the Earth sciences in Europe. It aims at facilitating the integrated use of data, models, and facilities, from both existing and new distributed pan European Research Infrastructures, allowing open access and transparent use of data.

The TCS MSL network allows researchers to collaborate with other labs and scientists. By becoming part of the rapidly growing TCS MSL network, new laboratories are offered a platform to showcase their research data output, laboratory equipment and information, and the opportunity to open laboratories to guest researchers through the Transnational Access (TNA) program.

The EPOS Multi-scale laboratories offer researchers a fully operational data publication chain tailored to the specific needs of laboratory research, from a bespoke metadata editor, through dedicated, (domain-specific) data repositories, to the MSL Portal showcasing these citable data publications. During this process the data publications are assigned with digital object identidiers (DOI), published with open licenses (e.g. CC BY 4.0) and described with standardized and machine-readable rich metadata (following the FAIR Principles to make research data *F*indable, *A*ccessible,

*I*nteroperable and *R*eusable. The TCS MSL is currently working on linking these data publications to the EPOS Central Portal[1], the main discovery and access point for European multi-disciplinary data, and on increasing the number of connected data repositories.

Keywords: Multi Scale Laboratories; Thematic Core Services; EPOS; MSL Subdomain

1. Introduction

In a world that demands increasingly interdisciplinary collaboration across scientific communities, solid Earth science laboratories are challenged with finding new forms of collaboration, sharing resources and access to laboratories, and developing best practices for laboratory routines as well as sharing data in standardized form.

The Multi-scale Laboratories bring together the scientific fields of analogue modeling, experimental rock and melt physics, paleomagnetism, geochemistry and microscopy in Europe as part of the European Plate Observing System[2] (EPOS). EPOS is a pan-European e-infrastructure framework with the goal of improving and facilitating the access, use, and re-use of solid Earth science data and research infrastructures [Cocco et al., 2022]. Today, the Multi-scale Laboratories (MSL) are one of the nine Thematic Core Services (TCS) of EPOS European Research Infrastructure Consortium (ERIC), which was established in 2019. The Multi-scale Laboratories represent a wide range of world-class laboratory infrastructures ranging from high pressure and temperature experimental facilities, to electron microscopy, micro-beam analysis, analogue modeling and paleomagnetic laboratories (Figure 1).

The main MSL activities range from 1) building a community network at the European level, 2) fostering transnational access to the partner laboratories, to 3) supporting open science by developing specific data services. Enhanced coordination and communication inside the European solid Earth science laboratories, complemented with services to increase curation of and access to laboratory data is needed to effectively contribute to solving the grand challenges of the society. The framework provided by the EPOS ERIC represents a unique opportunity to enact this change.

The largest effort for MSL has been the development of dedicated data services that can be used by the community to openly share their data in citable and reusable form. Until recently, most data produced by the various solid Earth laboratories and networks have been made available to the target scientific community only in printed form as final products (maps, tables or graphic plots). This severely limited the re-use and/or further elaboration of these, and especially the underlying raw data, in further research efforts. Consequently, much data remained inaccessible or restricted (i.e. accessible only on demand if you knew about their existence), often poorly documented and only stored on local computers without the perspective of long-term archiving. However, these data, produced in highly specialized experimental or laboratory environments, are unique and crucial to serving society's needs for geo-resources exploration and for geo-hazards mitigation. To model resource formation and system behavior during exploitation, we need an understanding from the molecular to the continental scale, based on experimental and analytical data.

Only recently, publishers started attempts to request access to data underlying published results and recommend researchers to use data repositories that are aligned with the FAIR Principles [Wilkinson et al., 2016], i.e. to make data *F*indable, *A*ccessible, *I*nteroperable and *R*eusable [e.g. COPDESS, Hanson et al., 2015, Stall et al., 2018]. However, those publisher's requests are generally not controlled rigorously and are often met by authors with a minimum of transparency and documentation of data, e.g. by providing uncurated data via generalist repository services. To improve the quality of published data, the EPOS MSL community decided to collaborate with appropriate domain repositories (e.g. GFZ Data Services) that provide additional curation services for data and metadata and focus on the development of tools and guidelines to increase the quality of metadata and data and, at the same time, reduce the workload for the individual researcher.

1 https://www.ics-c.epos-eu.org

2 https://epos-eu.org/

The largest challenge for MSL was the large variety of methods, instruments and data types used across the subdomains. E.g., the MSL infrastructures focus their research on Earth processes that encompass length scales ranging from the nano- and micrometer levels (electron microscopy and micro-beam analysis) to the scale of experiments on centimeters-sized samples, and to analogue model experiments simulating processes happening at the regional, to the tectonic plate, to the global scale (Figure 1). How much data homogenization is possible and how much diversity is required to keep the authenticity of each subdomain and at the same time develop the MSL community as the strong interdisciplinary network it represents today?

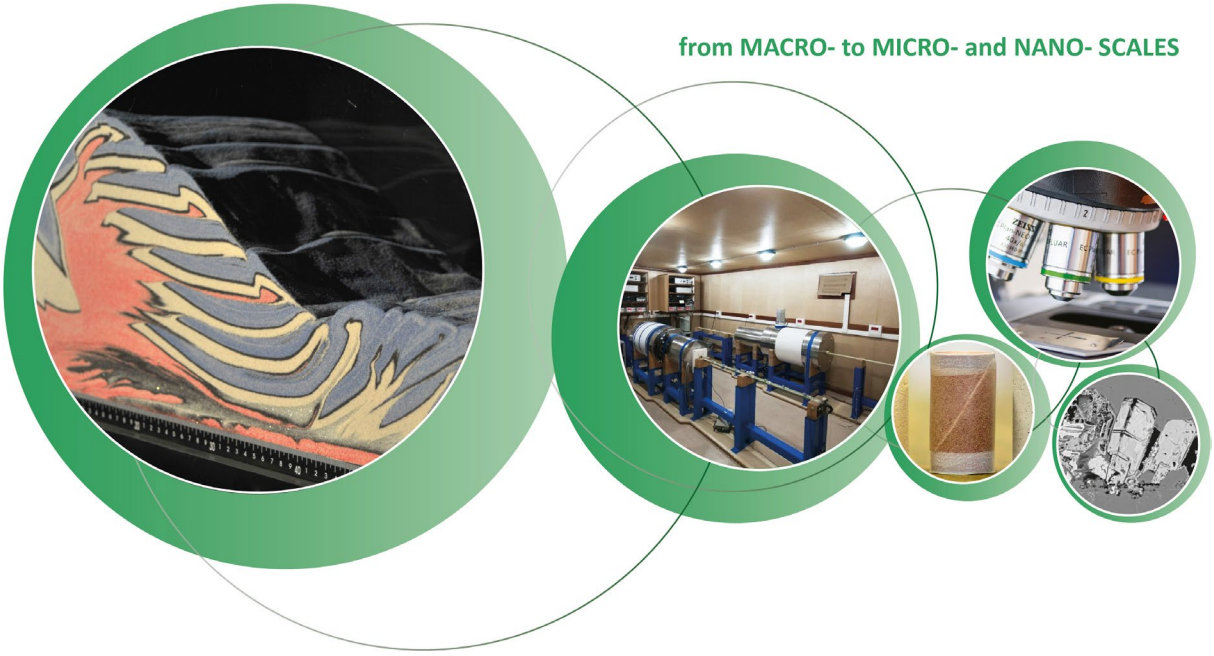

Figure 1. The Multi-scale Laboratories bring together the scientific fields of analogue modeling, experimental rock and melt physics, paleomagnetism, geochemistry and microscopy. The scientific focus ranges from tectonic to depositional processes, from hazards to geo-resources and from macro- to micro- and nano-scales.

2. The MSL Community

2.1 The History of MSL

The Multi-scale Laboratories started out as a new and very heterogeneous community of experimental and analytical laboratories ("Working Group 16") within the frame of the EPOS Preparatory Phase (EPOS-PP, 2010-2014, funded under the European Commission's Seventh Framework Programme). During this period the foundation was laid to turn small-scale infrastructures into a coherent, effective, and collaborative structure for scientists. As such the achievements of the Preparatory Phase were pivotal for the subsequent Implementation Phase (EPOS-IP, H2020-EU.1.4 Programme[3]) starting in 2015 when the TCS MSL was composed of ten laboratories in eight European countries, including consortia of laboratories like CNRS or NERC, bringing together the scientific fields of analogue modeling, paleomagnetism, experimental rock and melt physics, geochemistry, and microscopy. The network of the initial participating partners had four main aims: 1) to increase the coherence of the fragmented European experimental community, 2) to enhance the long-term sustainability of individual laboratories, 3) to strengthen existing, yet unofficial, collaborations and encourage new partnerships and synergies, and 4) to work towards the standardization and provision of data, data products and data services. During the five years of the

3 https://cordis.europa.eu/project/id/676564

EPOS Implementation phase, the TCS MSL has successfully created a network of European solid Earth science laboratories and scientists, which promotes and supports the exchange of valuable research data guided by the FAIR Principles ["*F*indable, *A*ccessible, *I*nteroperable, and *R*eusable", Wilkinson et al., 201]).

To build the MSL community, regular community meetings are held, both in correspondence with international conferences (mainly EGU General Assemblies) and independently as dedicated events (Utrecht 2013 and 2017, Potsdam 2016, Rome 2017, Barcelona 2018, Montpellier 2019). At these community meetings splinter meetings are held for the specific subdomains, together with sessions on specific topics and community-wide presentations. Such MSL community meetings have also promoted the launch of a Special Issue in the EGU journal Solid Earth on analogue modeling. MSL is regularly presented at major international conferences such as EGU, AGU, Goldschmidt, as well as smaller domain-specific conferences (e.g., GeoMod 2021, GeoMünster 2019, GeoUtrecht 2020, GeoKarlsruhe 2021). There have been several MSL workshops, including tutorials for data publications, which are open to the wider scientific community. In addition, in 2021, the MSL community started a seminar series on topics cross-linking the scientific subdomains.

Today, our network includes 86 laboratories from eleven countries (Figure 2). Until now, MSL has published 103 data sets via GFZ Data Services, the first repository open to all TCS MSL members from all subdomains. Three rounds of pilot TNA calls were successfully coordinated in 2017, 2018, and 2019. We have a dedicated community portal (MSL Portal) that provides access to the data and detailed information about the laboratory's infrastructure and equipment, and observe an increasing interest from solid Earth laboratories across Europe for joining our network.

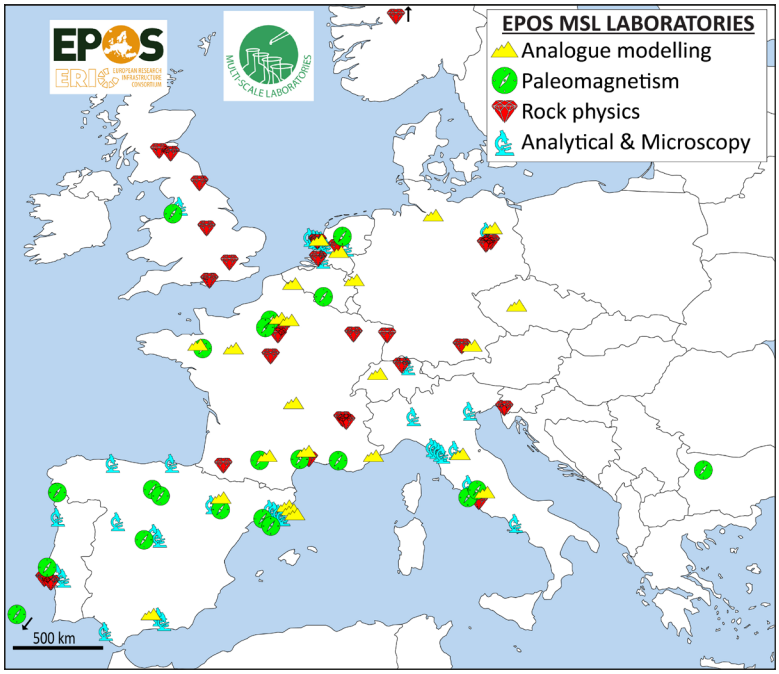

Figure 2. Geographical distribution of EPOS MSL Laboratories and their subdomains.

2.2 MSL Subdomain: Analogue modelling

Analogue modeling in Earth Sciences aims at understanding the physics/dynamics of a variety of geological processes operating at different spatial and temporal scales. The analogue community utilizes physical models scaled from a natural prototype (e.g. a mountain belt) down to the laboratory environment. Analogue modeling is a laboratory-based method and has been part of EPOS MSL since the beginning of the EPOS Preparatory Phase in 2010. Building on earlier initiatives to setup national and cross-national networks (e.g. TecMod F[4]), the initial

4 http://tecmodweb.free.fr/TecMOD_France/Accueil.html

EPOS network of laboratories already included several analogue laboratories in the Netherlands, Germany, Italy, and France. Since then, the analogue network quickly enlarged and strengthened. It now includes 22 laboratories from six countries and is expanding further, counting several declarations of interest per year to join the network.

Building a robust community network from scratch has been the most important success of the MSL analogue modeling team. The analogue community engages and empowers leading European scientists with the purpose of setting up activities that link methodological developments with scientific challenges and increasing the international leadership of the group beyond the European research arena, including transnational access. Researchers affiliated with the MSL analogue laboratories have run a dedicated recurring session at EGU, and recently organized a cycle of virtual/online/webinar seminars and community benchmark studies. Moreover, the community started to work on a "reference book" on analogue modeling and is organizing a thematic special issue in the EGU journal Solid Earth.

In the framework of the EPOS MSL, the analogue laboratories cyclically perform surveys of the experimental equipment available for the community at laboratory level, providing an up-to-date picture of the analogue modeling infrastructure landscape at the European level which is a key information allowing for an efficient scientific collaboration and technical interoperability.

The analogue community was first to publish their data with digital object identifiers (DOIs) and was a driving force in the development of the data description templates described below. Data sharing has become a tradition of the analogue community already early in the EPOS IP phase. As a benefit for the community there has been a rapidly emerging shared knowledge on properties and applicability of different analogue materials. The MSL Portal serves as a unique entry point to discover those contents by merging all information in the form of metadata harvested from the individual repositories.

2.3 MSL Subdomain: Paleomagnetism

The paleomagnetic community has been involved in EPOS since its early Preliminary Phase, being today a network of laboratories based in France, Italy, The Netherlands, Spain, UK, Poland, Belgium, and Bulgaria. MSL had a pivotal role in cataloging the instrumental equipment available at the laboratory level, providing a snapshot of paleomagnetic laboratories capabilities within EPOS and the international paleomagnetic community that is useful to achieve technical interoperability at European level as well as for organizing TNA.

Paleomagnetic data provision started with directional, magnetostratigraphic, magnetic susceptibility and paleointensity datasets, especially addressing the understanding of the evolution of sedimentary basins and for charting geo-hazard frequency. The next step will be the inclusion of environmental and rock magnetism data, following recent trends of paleomagnetic research and technology.

From the beginning on, the paleomagnetic MSL community decided to not develop a new paleomagnetic database for European data, but actively sought to establish an international collaboration with the Magnetics Information Consortium (MagIC)[5], the US-based global paleomagnetic database. This included, on one hand, the general agreement of MagIC to be open for data from European laboratories and joint workshops and tutorials encouraging researchers from European laboratories to publish their data via MagIC. On the other hand, it included the identification of metadata elements from MagIC and their harmonization with the MSL metadata schema as well as agreements on the metadata exchange format. From summer 2021 on, the first paleomagnetic data measured in MSL laboratories and published via the MagIC repository can be discovered and accessed via the MSL Portal.

2.4 MSL Subdomain: Rock- and Melt Physics

Rock- and Melt Physics is the subdomain for laboratories investigating the physical and mechanical properties of magma and rocks for understanding, modeling and forecasting their evolution across several space and time scales. It is one of the MSL subdomains included in all EPOS activities, including transnational access, since its Preliminary Phase and has been growing ever since.

5 https://www2.earthref.org/MagIC

Networking initiatives are mostly offered as dedicated workshops and conferences, either on an ad-hoc basis or recurring (e.g. during EuroConferences of Rock Physics and Rock Mechanics, or with topic divisions at AGU and EGU). There are also some professional networks through LinkedIn, as well as peer-lead ones (e.g. ROCKETH[6], Brit Rock[7]).

The subdomain's activities started with the identification of the laboratories and common vocabulary describing the data produced to form the metadata catalog and associated ontology. As most instruments are bespoke and/or home-made, this first ground of standardization has formed a significant step toward a greater understanding of commonalities for integrating FAIR data at the first stage of experimental planning. The common vocabulary developments caught the interest of the NSF-EarthCube project StraboSpot[8], an EarthCube[9] sponsored US initiative that is developing mobile and desktop applications to facilitate the collection and curation of field geology, microstructures, and experimental rock deformation data. StraboSpot was looking at integrating rock deformation laboratory data in their services. In early 2018, StraboSpot and the rock- and melt physics subdomain shared the MSL controlled vocabulary catalog and together built an international harmonization level for FAIR rock- and melt physics data.

2.5 MSL Subdomain: Geochemistry and Microscopy

Geochemical and microscopy laboratories are by number the largest and the most variable laboratories in the geosciences. There is a large range in techniques and equipment, and, consequently, a large variety in data and data types. Analytical techniques vary on the elements that can be analyzed with the respective detection limits. Analytical imaging based on microscopy can produce high-resolution images for the distribution of various elements, providing a prompt perception of mineral composition.

The Geochemistry and Microscopy Community has evolved since the EPOS Preparatory Phase. It now includes 24 laboratories for geochemistry, geochronology, mineral chemistry and fluid inclusion studies, providing compositional data from whole rock to micrometer or even nanometer inclusions (solid, liquid and/or gas) in minerals.

Due to the large variety within this subdomain, the standardization of data formats is extremely challenging, not only within the MSL community, but at a global scale. To join forces and harmonize at an international scale, EPOS MSL is an initial member of the upcoming OneGeochemistry Initiative[10] that is aiming at developing geochemical data standards at a global scale. Within OneGeochemistry, the EPOS Multi-scale Laboratories represent a European nucleus that shall bring perspectives and best practices from European laboratories to the global network and can help the MSL laboratories to implement the newly developed data standards.

Similar to the MSL Rock- and melt physics community, the geochemistry and microscopy community of MSL is currently developing a vocabulary for microscopy and microstructures in collaboration with StraboSpot. StraboSpot is developing an application, StraboMicro (Newman et al., 2021), for image management and data reporting, and since 2020 the MSL microscopy community is involved in vocabulary expansion and harmonization.

3. MSL Governance and Coordination

The MSL Governance and Coordination structure is pivotal for internal and external communication, further developing the MSL network, and any kind of outreach activities (Figure 3). The MSL Governance and Coordination is based on a consortium agreement, which is currently signed by eleven research institutes, located in eight countries. The Consortium Board acts as the decision-making body, which is composed of one representative per consortium partner. The Executive Committee is responsible for the day-to-day management of the activities of MSL. In addition, the Executive Committee is the link between EPOS ERIC and the Thematic Core Service MSL.

6 https://rockphysics.org/

7 http://britrock.org/

8 https://www.strabospot.org/

9 https://www.earthcube.org/

10 https://www.earthchem.org/communities/onegeochemistry/

The TNA Committee is responsible for the management of transnational access calls and interaction with the TNA providing laboratories. The Data Provider Committee is responsible for the data services and the data policy of MSL, and provides expert advice on the domain-specific vocabularies and metadata. The Consortium Board and the various committees of MSL have regular teleconferences to streamline their activities.

The mission of the EPOS TCS MSL is to create a unique entry point for collaboration and exchange through:

1) the organization of a **coherent and collaborative network** of European solid Earth science laboratories;
2) the implementation of dedicated **data services** that will allow to collect, preserve and share available and emerging laboratory data on the properties and processes controlling rock/geological system behavior at multiple scales for re-usability and interoperability with other solid Earth science data;
3) the development of a **Transnational Access (TNA) program** that will increase European state-of-the-art solid Earth science laboratories' attractiveness for researchers and contribute to increased researcher's mobility, cooperation and exchange.

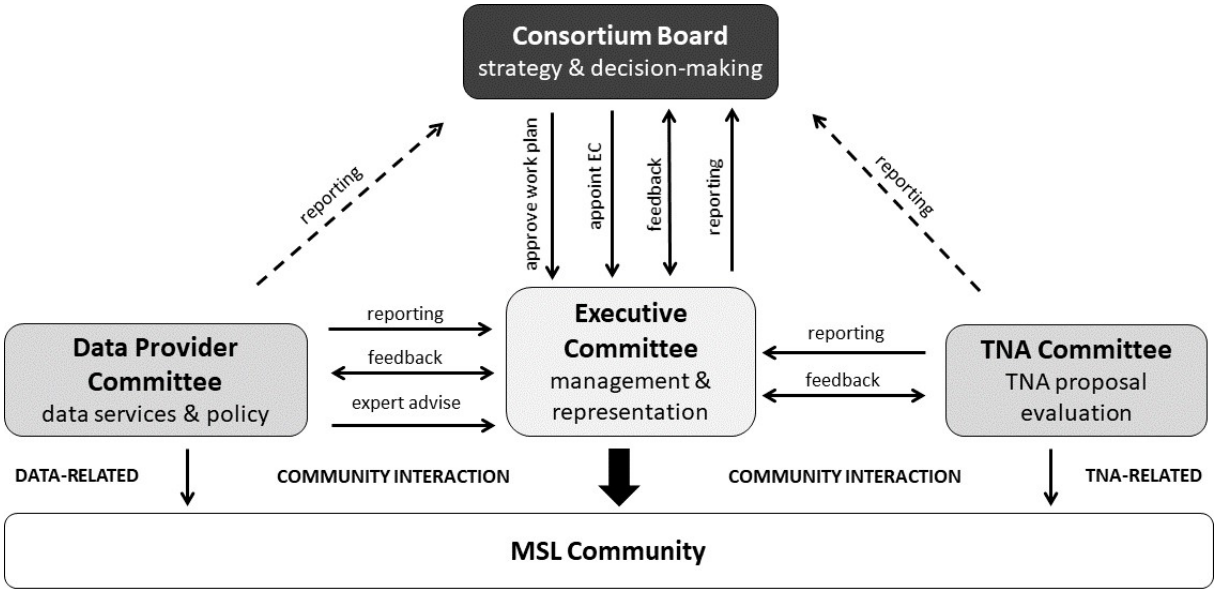

Figure 3. The MSL Governance and Coordination structure.

4. Multi-scale laboratories Transnational Access

Transnational Access (TNA) refers to the virtuous program allowing for a free-of-charge physical (i.e. access to laboratories in person) or remote (i.e., contracted sample testing) admission to, interaction with, and use of research infrastructures (here laboratories) for researchers in a country foreign to their own affiliation. With this strategy, the TNA program is fostering new opportunities for synergy, collaboration, and innovation in research (see also Wessels et al, 2022). TNA can be considered pivotal to experimental research. The use of different and often complementary experimental and analytical facilities as well as the exchange of laboratory practices and methodologies are key to the scientific aim of a research project. In particular, the opportunity of combining experimental and analytical techniques in the framework of the same study has become a prerequisite of many branches in the Earth Sciences. The TNA program of the MSL is at the core of this interdisciplinary approach and facilitates researchers in their scientific tasks by providing access to state-of-the-art laboratory infrastructures as well as fostering the generation of new standardized and FAIR datasets.

The Multi-scale Laboratories successfully developed and tested the TNA experience, with physical or remote access to participating facilities from all subdomains offered during three pilot calls (in 2017, 2018, and 2019) with the aim of set up, check, and refine the TNA procedures for the whole EPOS community, also in the perspective of

the service validation. The MSL TNA service has been provided through a call-for-proposals mechanism, setup by the MSL following the consolidated European TNA rules[11].

The number of available TNA providers increased progressively (5 in 2017, 22 in 2018 and 38 in 2019) and the numbers of users followed the same trend, with a success rate of 5/7, 23/28 and 25/25 approved over submitted proposals in 2017, 2018 and 2019, respectively. This statistical data highlights the organizational strength of MSL and the interest in the initiative demonstrated by both the TNA laboratories and the visiting researchers.

The user feedback related to the MSL TNA pilot experiences showed a straightforward correlation between the availability of financial support and the number of applications to specific facilities. MSL supplier laboratories provided free-of-charge access to their facilities and, in some cases, supported the travel costs (i.e., accommodation and sustenance) thanks to national funding (e.g., the Netherlands) and/or institutional funds (both of users and providers). From all laboratories participating in the second call, 22% of the laboratories were able to offer full financial support (including travel and accommodation) to the users. These received 46% of the applications proving the attractiveness of fully supported research opportunities.

The user feedback reports have shown the importance of financial support from research infrastructures to TNA Users in making the TNA a success. Only 5% of the TNA Users reported that they would have been able to carry out TNA research activities without any financial support. These numbers show the requirement and importance of EPOS to implement a stable and clearly defined TNA Program for fund provision and provide a steady financial support to this pivotal initiative.

A successful story of the TNA pilot cases is represented by the publication of the first green and gold open access articles [e.g., Ge et al., 2019a,b; Michail et al., 2021; von der Linden et al., 2021] and datasets collected during the TNA experiences within EPOS TNA framework, which proved the success of TNA General Principles with respect to the application of EPOS data policy in the TNA cases [e.g., Beekman et al., 2019; Deng et al., 2018; Román-Berdiel et al., 2019].

Overall, the MSL laboratories that participated in TNA made very good experiences and see this option as an important measure to strengthen international scientific collaboration. These TNA activities, however, were fully financially supported by the participating laboratories or national funding agencies. To further expand TNA on an EPOS-wide level, however, additional funding is required. A summary of the MSL TNA pilot calls and results are also described in Wessels et al. [2022].

5. Working on data services – open data for MSL

Data from Multi-scale Laboratories can be described as "long-tail" of research data [Heidorn, 2008]. These data are small in terms of memory size but highly variable, hence difficult to standardize and curate. But they represent a large portion of the total quantity of research data. Figure 4 provides an overview on the different data types from MSL laboratories.

Following the FAIR principles [Wilkinson et al., 2016], research data should be Findable, Accessible, Interoperable and Reusable. Publishing data and other research output under these principles requires the generation of machine-readable metadata and the use of persistent identifiers for uniquely identifying people, institutions and funders, as well as for cross-linking the published data with research articles, related software or physical samples that were analyzed to derive the data. To facilitate data discovery in portals, research data should be indexed with domain keywords.

To include data from European laboratories in the EPOS infrastructure, we have agreed on several strategies: 1) DOI-referenced data publications via domain repositories have been identified as best practice for FAIR sharing data, especially in long-tail communities with highly variable data types: researchers are describing their data via metadata editors that convert the information to standardized and machine-readable metadata formats (XML, JSON), These are made available via standard application programming interfaces and/ or embedded in the websites (Schema.org) by the repository. Data curation by domain scientists and additional technical description via data description templates further improve the quality of the published data and their documentation. 2) We en-

11 https://ec.europa.eu/info/files/european-charter-access-research-infrastructures_en

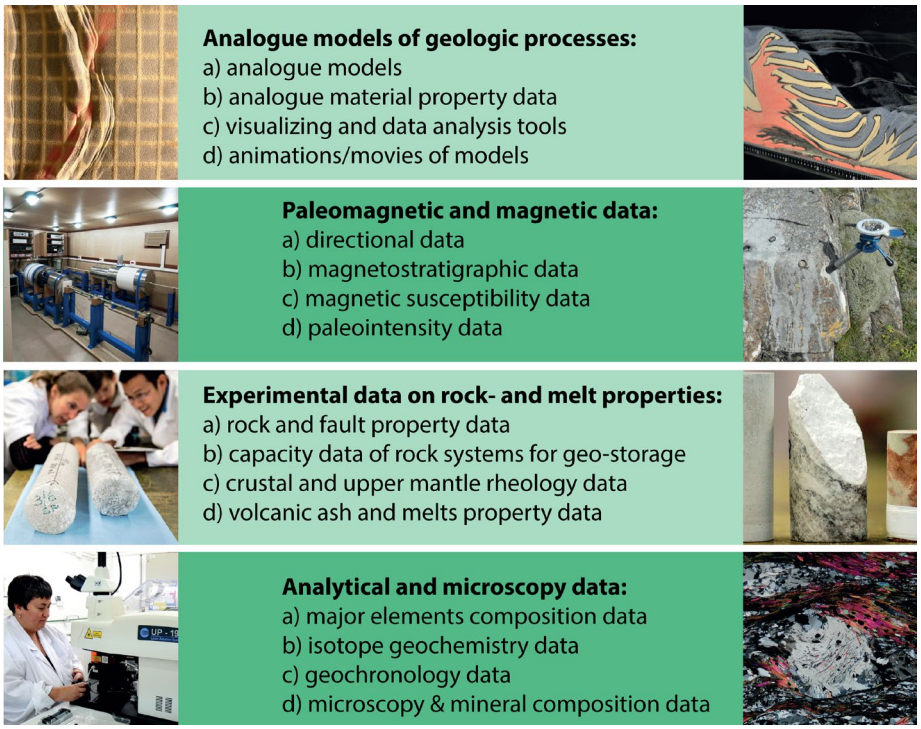

Figure 4. Typical data types for MSL subdomains

rich the DOI metadata with a community-agreed metadata schema that is based on the combination of several controlled linked-data vocabularies used in the geoscience domain.

In the following sections we describe the actual status of our data services as developed during the last seven years. New developments, like the ROR identifier (ROR = Research Organization Registry, a globally unique and resolvable identifier for institutions) or persistent identifier for instruments, but also the revision of our controlled vocabularies and identification of new vocabularies to be included in our services are currently in progress and discussed in the Outreach and future perspectives for MSL part below.

The main partner repository of MSL to date is GFZ Data Services[12], a domain repository for geosciences, hosted at the Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences[13]. GFZ Data Services has a focus on the curation of long-tail data. The repository implements the FAIR principles by 1) the provision of comprehensive domain-specific metadata that integrate controlled "linked-data" domain vocabularies following international metadata standards for data discovery (DataCite, ISO19115); 2) complementing the metadata with detailed data descriptions or reports; and 3) embedding the research data in wider context by providing cross-references through persistent identifiers to related research products and people or institutions involved (DOI, IGSN, ORCID, Fundref). The visibility and discovery of the data is established by the registration of digital object identifiers (DOI) with DataCite[14] and the dissemination of metadata in standard protocols. To facilitate discovery through internet search engines like Google, all DOI landing pages of GFZ Data Services embed machine-readable metadata following Schema.org and present xml versions of the metadata in DataCite and ISO 19115 standards. The OAI-PMH interface can be queried with search parameters which allows the harvesting of a specific subset. In addition, GFZ Data Services feed links of data and related research products into Scholix[15], which allows direct linking of data publications and scholarly literature, even when the data are published years after the article.

12 https://dataservices.gfz-potsdam.de/

13 https://www.gfz-potsdam.de/

14 https://datacite.org/

15 http://www.scholix.org/

5.1 Development of MSL-specific metadata schemas and vocabularies

The first step in the data publication chain was the development of a community-agreed metadata schema that describes the different MSL data types as comprehensively as possible. This was pursued by the integration of controlled vocabularies in existing metadata standards for data discovery (DataCite, ISO19115) used by the partner repositories. These vocabularies describe, for example, the geologic setting, geologic and geomorphic features, observed processes, the geochronological time scale, lithology and other material properties, laboratory instruments, monitoring devices and software used to process the data. Controlled vocabularies are provided as nominal lists in the EPOS MSL metadata editor. Whenever possible, MSL used registered, OGC compliant, controlled vocabularies that are governed by international groups, e.g. IUGS CGI geoSciML[16], INSPIRE GEMET Keywords, NASA Global Change Master Directory (GCMD), International Chronostratigraphic Chart, and provided in RDF format. Only for specific fields, like rock analogue materials or custom-made devices and instruments, we developed individual vocabulary lists that shall be also provided in RDF format and governed by the EPOS MSL community in the future. MSL researchers are further encouraged to include their ORCIDs in the metadata, and to cite related articles, datasets, software or samples by their related identifiers (using DOI, IGSN[17], URLs). We further agreed on keyword lists for the MSL subdomains and data types used for all data publications as well as for the laboratory descriptions described below.

Moreover, the name of the laboratories is included in the DataCite metadata (as "contributor: hosting institution") in standardized form from a newly created list that is maintained at GitHub by the MSL community[18]. The implementation of the ROR Identifier for the laboratories institution in this list is in preparation by GFZ Data Services. These stategies enable direct links of the data with the originating laboratory as well as easy implementation of search facets in the MSL Portal. Moreover, as the same vocabularies were also used for the description of the laboratories (i.e. instruments, measurement types, materials), direct links between data and instruments, measurement and materials can be implemented in the MSL Portal.

5.2 The MSL Metadata Editor and Data description templates

Software solutions, like online metadata editors, support scientists in generating metadata in machine actionable form and thus bridge the scientific with the data curation world. For their data and software publication activities, GFZ Data Services provides an XML metadata editor[19] that assists scientists to create metadata in different international metadata schemas (ISO19115, DataCite), while being at the same time usable by and understandable for the scientists [Ulbricht et al., 2017, 2020]. The editor is publicly available on the internet without registration; a copy of the metadata (in XML format) can be saved to and loaded from the local hard disk; and scientists are not requested to provide information that may be generated automatically. To further improve usability, we offer a facility to search structured vocabulary lists, and an interactive geographic mapping tool that provides direct visual control on the geographic coordinates potentially relevant for the data (e.g. sample location, study area). As the main MSL partner repository, GFZ Data Services extended their metadata schema to fully include the community-agreed metadata schema describing MSL laboratory data, and developed an MSL-specific instance of their metadata editor[20] [Ulbricht et al., 2017].

Even with a very rich metadata schema, published data are often not sufficiently described for being reusable. Additional technical descriptions, like the definition of column headers of tabular data, the description of the experiment setup, documentation of software tools used to manipulate the data or an overview of a complex file structure or specific data format, are necessary to fully understand the data. We have further observed that many researchers are unaware what a data publication represents and especially what to include in a data description. This observation, together with the aim to further standardize MSL data, led to the development of data description templates. Originally initiated within the MSL analogue modeling community, the templates provide a framework for data

16 http://geosciml.org/resource

17 IGSN = International Generic Sample Number, https://www.igsn.org

18 https://github.com/ulbricht/pmdmeta/tree/eposlabs/resources/thesaurus/eposlabnames

19 https://dataservices.gfz-potsdam.de/panmetaworks/metaedit/

20 https://dataservices.gfz-potsdam.de/msl/

descriptions and are now used for all publications via GFZ Data Services. Data descriptions with the templates are fully harmonized, curated by the repository staff and provided as a PDF data description in the data download folder.

The "Data Description Templates for data publications with GFZ Data Services"[21] are publicly available in a commented version and a blank version with the headlines and header only (see Figure 5). Using the templates provides additional standardization to our data publications and increases the metadata quality. Their uniform format 1) aids in comprehension by the reader, and 2) the use of the templates significantly reduces the curation work by repository staff because the researchers are providing comprehensive descriptions from the beginning on.

Figure 5. Example for data descriptions of GFZ Data Services for a data publication by Rudolf et al. (2021), The header (to the left) is followed by eight pages according to headlines of the template (to the right).

5.3 Collection of structured laboratory descriptions for EPOS

One precondition to enable transnational access (TNA) is to have detailed information on the laboratory infrastructure offering TNA. What instruments are available in which laboratory? Where are the laboratories located? Who is the laboratory manager? How are the TNA access rules?

To enable this, we have developed a structured metadata schema for collecting this information. A subset of this description is currently provided In the EPOS TNA portal[22] and in the MSL Portal[23]. The full information will be made available in the MSL Portal with the next release.

21 https://gfzpublic.gfz-potsdam.de/pubman/item/item_5007103

22 https://epos-no.uib.no/epos-tna/facilities

23 https://epos-msl.uu.nl/

The general strategy for the development of the laboratory metadata schema was to enable direct links of laboratories with data derived from them or to search for all data measured with a specific instrument. In addition, it was required to identify laboratories that have a specific instrument during a TNA application process. To achieve these tasks, we used the same controlled vocabulary that we developed for describing data publications, including a controlled list of laboratory names and hosting institutions managed by MSL via GitHub[24].

The laboratory description schema is available in Excel format with specific controlled-vocabularies for each subdomain. Whenever possible, complex nominal lists were included in the Excel sheets to avoid spelling errors. Table 1 shows the general structure of the laboratory descriptions. The hierarchical vocabularies underlying laboratory services are specific for each subdomain and provided as nominal lists in the templates.

General information for all subdomains		
Laboratory Information	**Summary information on the laboratory, including specification and instrumentation: this part is also used for the general description of the laboratories in the MSL Portal (free text, max 4000 characters)**	
Practical Info about the facility	**RI name** (= institute, consortium running the laboratory), *type* (= laboratory), ID, **name, address, geographic coordinates**, website, *research field* (MSL subdomain), **contact person** (**name**, ORCID, **email, affiliation**), **TNA participation**	
Laboratory services	Equipment	**type***, **group***, **name***, **brand**, **contact person**, email, website, specific and comments, quantity, references
	Measurement	**type***, **group***, **name***, **brand**, website, type specifics and comments, references

Table 1. General schema for MSL laboratory descriptions. Bold fields are mandatory information, bold italic fields are prefilled, asterixis mark fields with nominal lists (see also Figure 4). The laboratory services are specific for each subdomain.

Figure 6 illustrates the richness of the hierarchical vocabulary for an example from the Paleomagnetism Community: Demagnetizer and magnetizer are one of five different general methods in which paleomagnetic instruments are used (i.e., controlled-field area, demagnetizer, field samples, rock magnetism instrument and rock magnetometer). Demagnetizers and magnetizers themselves can be subdivided in five different Demagnetizer and magnetizer types (= Equipment Group), each with 1-8 different instruments (= Equipment Name). Instruments not included in the list can be added by selecting "other" and providing the name of the instrument. After approbation by the MSL paleomagnetic community, the new instrument will be included in the vocabulary list with the release of the next version. Similar hierarchical lists exist for measurement types, materials and instruments for all MSL subdomains. New laboratories with interest to join MSL are requested to provide their laboratory descriptions with the templates before they are officially accepted as part of the MSL community.

5.4 MSL Portal

The EPOS MSL Portal is the central access point for the MSL community. Users can find information on the participating laboratories, the subdomains, and all data published by different data repositories. Based on the community-agreed rich metadata schema, the portal offers search facilities by which a user can directly link the data with the laboratory they were measured at, can find data for specific instruments, and can get an overview of available domain-specific data.

24 https://git.science.uu.nl/o.a.lange/inframsl/-/blob/master/labnames.json

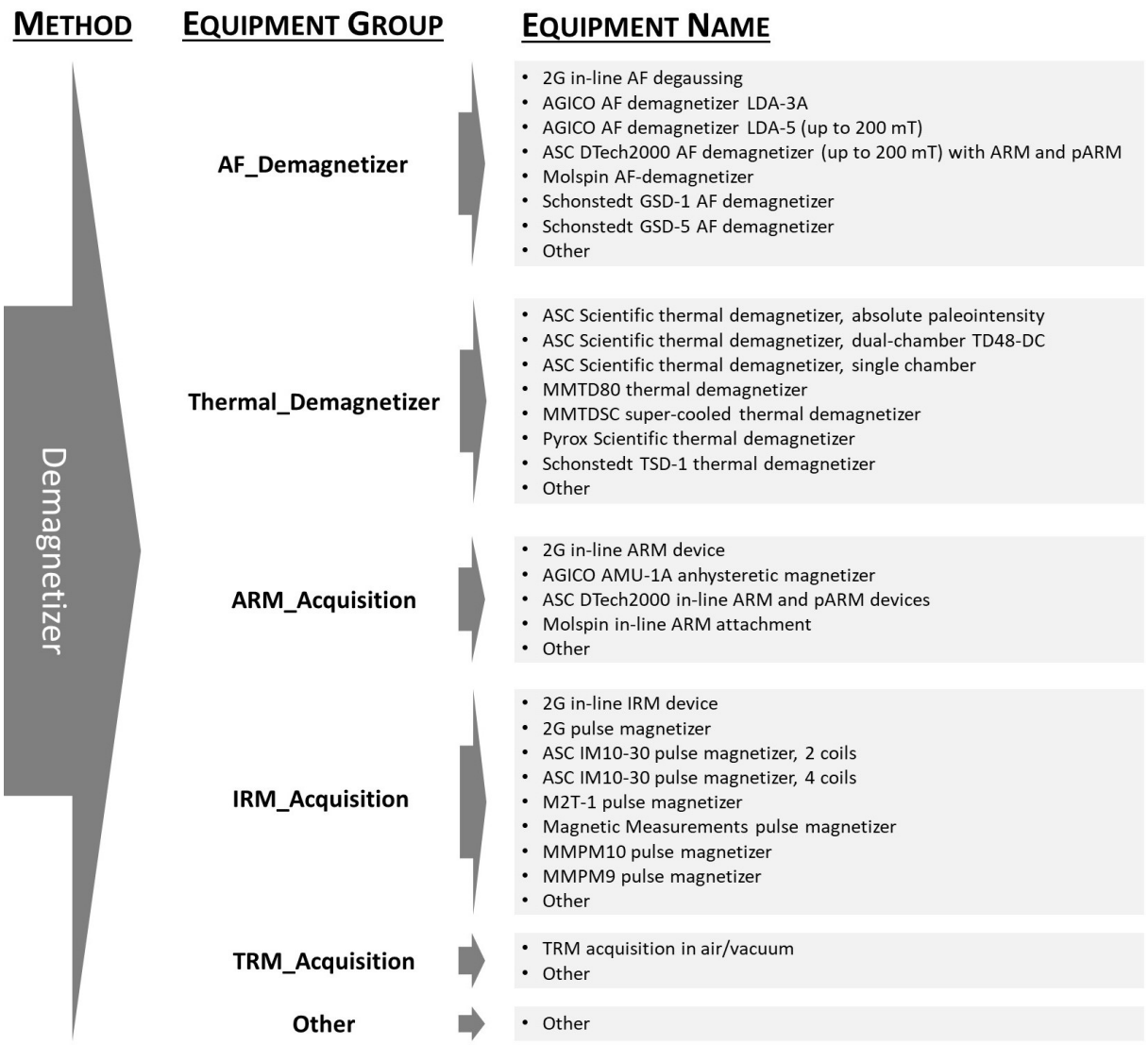

Figure 6. Example of the controlled vocabulary describing demagnetizer and magnetizer instruments for paleomagnetic measurements. This hierarchy is implemented as nominal lists in the laboratory description templates.

The portal runs on the CKAN software for data catalogs. It is populated with metadata that are collected from the various repositories where MSL researchers may have published their data (see Figure 7). The MSL Portal provides a node for the further dissemination of MSL metadata through its connection with the EPOS Central Portal.

One of the biggest challenges for the development of the MSL Portal is the harmonization of metadata from different repositories. While data published via GFZ Data Services includes the full MSL metadata schema, other data, especially when published via institutional repositories serving more than one scientific domain (e.g. DIGITAL.CSIC in Spain) or harvested from existing global databases (e.g. MagIC[25]) can only partly adopt the MSL metadata schema. This challenge is met by a metadata enrichment process that automatically applies during the harvesting process such that the MSL community standards are met as far as possible. To facilitate this, MSL researchers are encouraged to provide as many keywords from the MSL metadata schema as possible as free keywords in the DataCite metadata used by all repositories assigning DOI. In addition, we have agreed with DIGITAL.CSIC and MagIC to include the same laboratory names from the MSL laboratory list on Github[26].

25 https://www2.earthref.org/MagIC

26 https://github.com/ulbricht/pmdmeta/tree/eposlabs/resources/thesaurus/eposlabnames

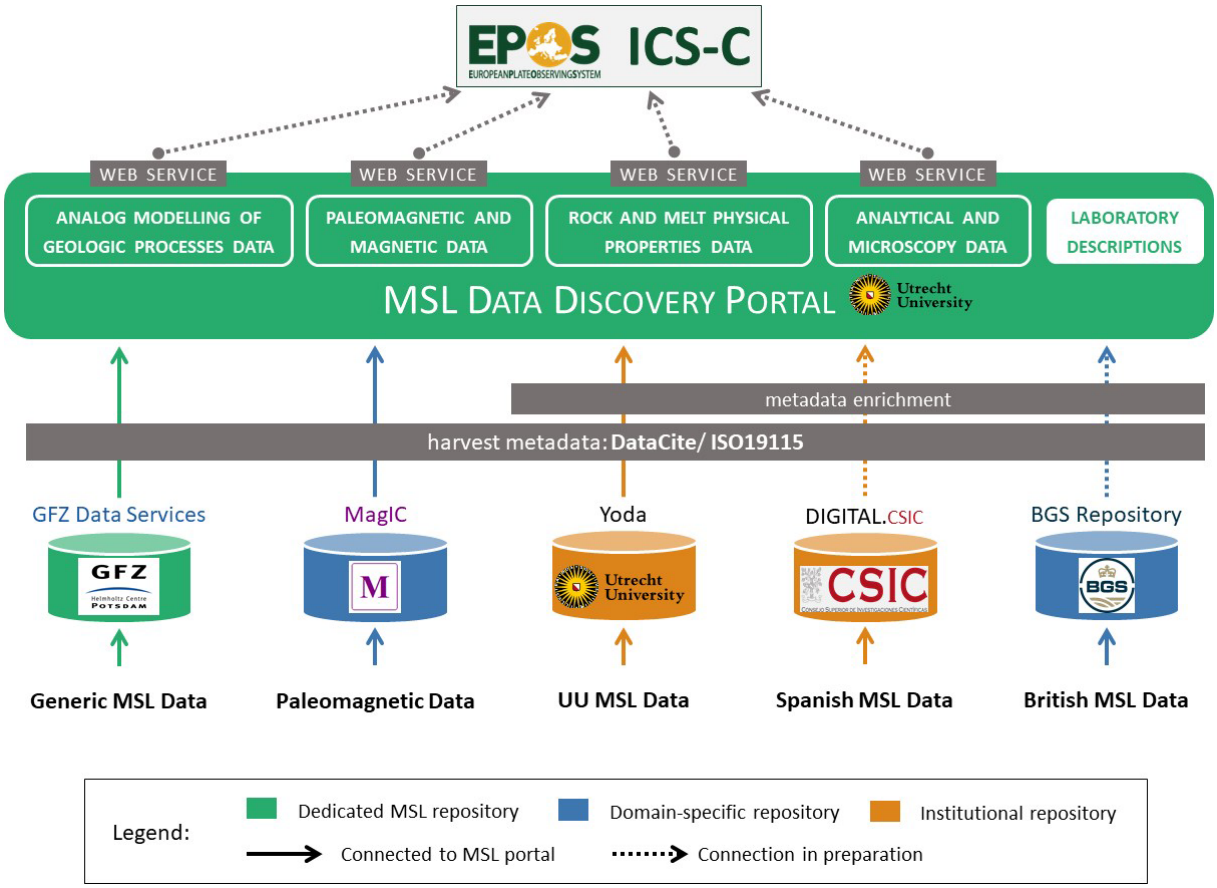

Figure 7. Schematic overview of current and planned repositories linked to the TCS MSL Portal. GFZ Data Services (green) offers data publication services for all subdomain data with the most complete metadata. MagIC is the global database and repository for paleomagnetic data. Yoda is the Utrecht University repository and the third repository connected to the MSL Portal. DIGITAL.CSIC and BGS represent the Spanish and British data repositories, respectively. These repositories are planned to be linked to the portal in the near future.

What we see is that currently only few repositories that are relevant for the MSL community can provide their metadata in full compliance with the MSL metadata structure. This is a common challenge when collecting metadata from a heterogeneous landscape of infrastructures. The MSL community roadmap is aimed at increasing metadata harmonization, thereby fostering the consistent use of the MSL metadata schema. In the meantime, i.e. the enrichment processes in the MSL Portal take care of an increase in findability, not only within the portal itself, but also at the level of the EPOS Central Portal.

6. The MSL Data publication chain

The MSL communities have agreed to not develop community specific data bases or repositories for European laboratory data, but seek synergies with existing domain repositories and best practices, like DOI-referenced data publications, for making their data available. These FAIR-aligned repositories provide workflows for automated metadata exchange and use internationally agreed metadata standards and protocols that are provided in machine-readable form. Key research infrastructure is the specifically developed MSL Portal. This portal aims at collecting all metadata provided by the different repositories and makes them findable and accessible in a broader context of pan-European laboratories. In addition, it provides information on MSL laboratories and enables specific search options for data, equipment, laboratories, etc., based on the MSL metadata schema.

The MSL data publication chain enables researchers from the MSL community to share their data through data publications, using harmonized rich metadata, combined with information on the laboratories where the data was

generated. The data can be published via different channels: via the dedicated MSL repository GFZ Data Services or via institutional or domain repositories that are using DOI for their data publications. The associated metadata are harvested (and enriched whenever necessary) by the MSL Portal which further makes them available to the EPOS Central Portal.

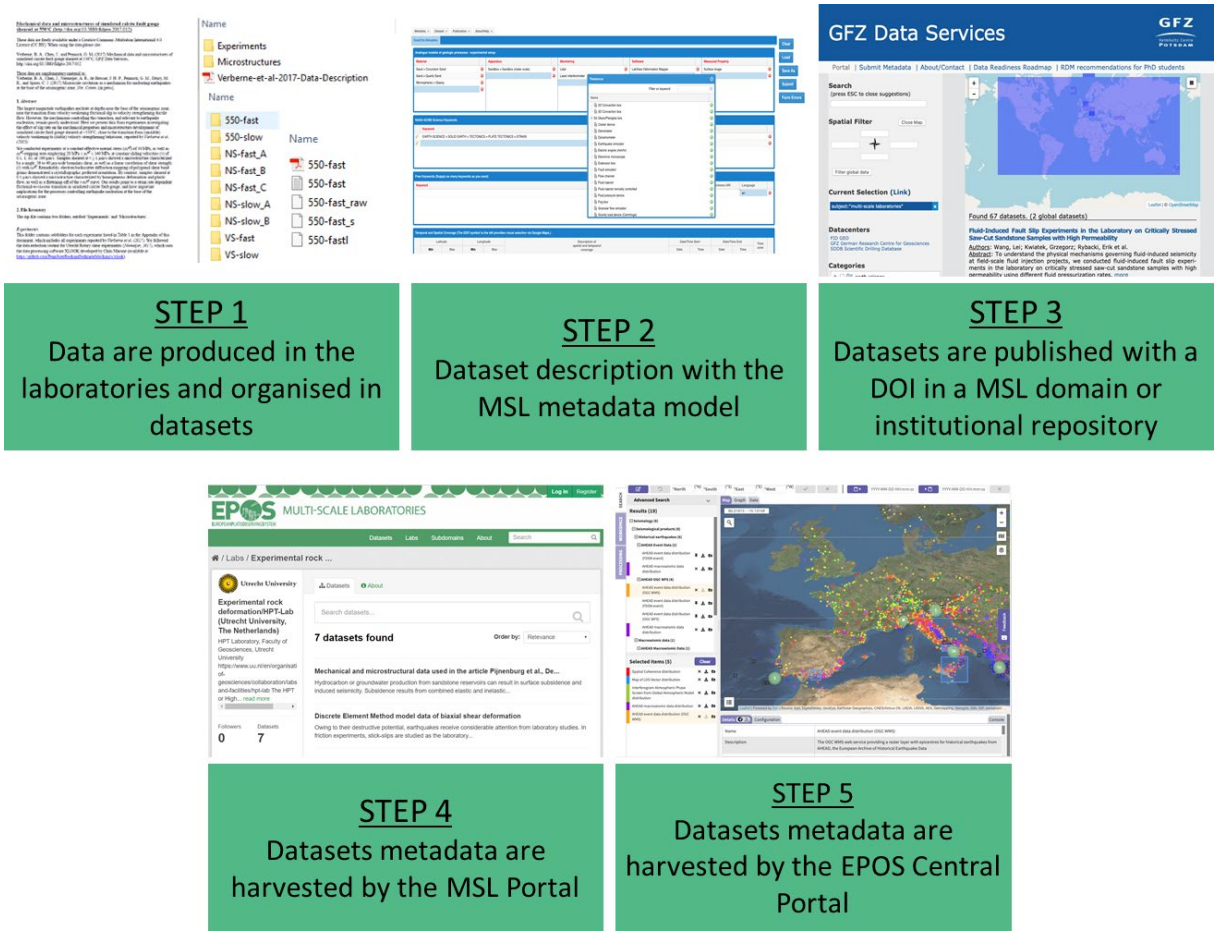

Figure 8. Schematic illustration of the steps of the MSL data publication chain.

The data publication chain offered to the MSL community can be described in five steps (Figure 8):

1) Data are produced in the laboratories and organized into datasets.
2) Dataset description (by the researcher) is done with the help of the MSL Metadata Editor and the data description templates.
3) Datasets are published with a DOI in the dedicated MSL repository GFZ Data Services or in other (institutional or domain) repositories. This entails the conversion to machine-actionable standardized metadata (via the metadata editor), data and metadata curation, and assignment of a DOI to the data publication. GFZ Data Services is available for publication of data for all subdomains, which allows for the richest addition of metadata via the MSL Metadata Editor. The BGS repository is available for UK rock physics data. Paleomagnetic data can be published via the domain repository MagIC. Spanish data can be published via Digital.CSIC[27], data from Utrecht University via Yoda[28].

27 https://digital.csic.es/

28 https://www.uu.nl/en/research/yoda

4) Dataset's metadata from the various repositories are harvested in the MSL Portal. Metadata are harvested via a standardized API and enriched (if required and possible). Figure 7 shows the repositories currently connected to the portal and those who are planned to be connected.
5) The metadata from data publications and the laboratory descriptions in the MSL Portal will be integrated in the EPOS ICS-C portal via web services.

7. Outreach and future perspectives for MSL

The MSL community is constantly expanding, uniting a growing community of laboratories in Europe following the strategic aim to become the pan-European network of solid Earth laboratories. After the ending of the EPOS-IP phase, ten new laboratories have requested to join the Multi-scale Laboratories. In order to guarantee a high quality of laboratories joining MSL, an ingestion procedure was developed. New laboratories have to provide an application letter including a motivational letter and the description of the laboratory and equipment via the MSL laboratory description templates (see section 5.3). The MSL Consortium Board reviews the application and has the ultimate say in approving the new laboratory as part of MSL. All laboratories that are in the European solid Earth domain who are willing to share their data are welcome to join the MSL Community.

In 2021, four laboratories from the geo-energy community requested to join MSL. Three of these large-scale laboratories already were part of the candidate TCS Geo-energy test beds in EPOS-IP. The MSL Consortium Board agreed on the request and their inclusion as a new MSL subdomain has started in early 2022.

An extensive revision of the controlled vocabularies already used, and the integration of new vocabularies and facets for new subdomains in the MSL Metadata Editor and the MSL Portal is ongoing. We focus on linked-data vocabularies in SKOS/RDF format that are maintained by a community, e.g. the International Union for Geological Sciences (IUGC). The rapid progress in research data management internationally include the development of more vocabulary services, like GeoERA GIP-P[29], that will be explored during upcoming revisions, especially for the new geo energy subdomain. The revision/expansion of the rock physics vocabulary and the new development of a microscopy vocabulary is pursued in close collaboration with the EarthCube initiative StraboSpot[30] and the H2020 project EXCITE[31].

The further development of the MSL metadata schema will be closely aligned with the ongoing developments in the international data curation landscape and recommendations by groups, like the Research Data Alliance (RDA). MSL is encouraging repositories to, e.g., include the ROR identifier in their metadata and laboratories to consider using the newly developed persistent identifier for instruments.

On the technical part, there are two major ongoing developments: the expansion of the MSL Portal, and the development of a Facility Access System (FAST) supporting TNA activities. The current MSL Portal was set up with two major functionalities in mind: 1) collecting metadata from multiple repositories, and 2) sharing these metadata in a format compliant with the EPOS Central Portal via dedicated web services. Although end users can easily find data publications within the current set up, upcoming portal developments will support more options to share information coming from the laboratories (description of laboratories and equipment) and the community as a whole. These functionalities must enforce the role of the portal as the central point of access for important MSL-related content and updates.

The EPOS TCS Multi-scale Laboratories have executed three successful TNA pilot calls. During these calls, the MSL TNA procedures and policies were designed, tested, and refined. Call-related information was collected by TNA coordinators using spreadsheet software, which was a laborious and error-prone process. To optimize this process in future TNA calls, the MSL have decided to develop the dedicated Facility Access SysTem (FAST), which is funded by the Dutch National EPOS Infrastructure EPOS-NL and the H2020 project EXCITE (Wessels et al., 2022). FAST consists of two main components: 1) a database, which stores information about facilities and call-related information, and 2) a coordination module, which regulates the coordination, application, and review procedures. FAST will be harvestable by the EPOS TNA Brokering System, enabling the dissemination of TNA-related metadata

29 https://geoera-gip.github.io/about/

30 https://www.strabospot.org/

31 http://www.excite-project.eu/

at the European level. We envision making FAST available to all TCS providing TNA in the future (Wessels et al., 2022).

Reaching beyond the community, MSL is collaborating with several international initiatives and frameworks. Together with the AuScope Geochemistry Network[32], EarthChem[33] and Georoc[34], MSL is part of the OneGeochemistry Initiative[35], a global geochemical data network that facilitates discovery and access of standardized geochemical data through coordination and collaboration among international geochemical data providers. This network is currently being established.

8. Conclusion

In this article the various aspects of the mission of the Multi-scale Laboratories are set out. MSL is a coherent and collaborative network of European solid Earth science laboratories, uniting researchers in the four subdomains: Analogue Modeling, Paleomagnetism, Rock and Melt Physics, and Geochemistry and Microscopy. The building of a community relies heavily on the possibility to interact with each other. Due to the global pandemic, this proved challenging in the past two years, and all meetings were moved to a virtual environment. The community showed to be flexible and resilient, and new initiatives such as the development of a special issue on analogue modeling of basins and an online seminar series were started during these online events.

Within MSL, dedicated data services have been developed, resulting in a tailored data publication chain that enables the MSL community to collect, preserve and share laboratory data for re-usability and interoperability with other solid Earth science data. The establishment of a Transnational Access program allows researchers access to European state-of-the-art solid Earth science laboratories.

In collaboration with StraboSpot new vocabularies for microscopy and microstructures are currently under development, as well as improvement of the existing rock physics vocabularies. MSL has collaborated with the Magnetics Information Consortium (MagIC) to allow harvesting of data publications by MSL researchers from the MagIC Repository to the MSL Portal.

After more than 10 years, MSL is a growing network, which is open for new laboratories to join. The community is actively integrated in international activities and thereby posing as a major player in the data curation landscape. MSL is aiming at fostering new opportunities for synergy, collaboration, and innovation in research and will continue to do so in the future.

References

Beekman, F., E. Willingshofer, D. Sokoutis, E. L. Pueyo, A. M. Casas, A. Peiro, T. Román-Berdiel, R. Soto, J.-M. Schönebeck, M. Warsitzka, and M. Rosenau (2019). Ring shear test data of iron powder – quartz sand mixture: a new marker material for analog modelling in a CT scanner at Utrecht University (The Netherlands) (EPOS TNA call 2017). GFZ Data Services, https://doi.org/10.5880/figeo.2019.018.

Cocco, M., C. Freda, K. Atakan, D. Bailo, K. Saleh-Contell, O. Lange and J. Michalek. (2022). The EPOS Research Infrastructure: a federated approach to integrate solid Earth science data and services. Ann. Geophys. 65, 2, https://doi.org/10.4401/ag-8756.

Deng, B., J. Schönebeck, M. Warsitzka, and M. Rosenau (2018). Ring-shear test data of different quartz sands and glass beads used for analogue experiments in the experimental laboratory of the Chengdu University of Technology (EPOS Transnational Access Call 2017). GFZ Data Services, https://doi.org/10.5880/GFZ.4.1.2018.003.

Ge, Z., M. Rosenau, M. Warsitzka and R. L. Gawthorpe (2019). Overprinting translational domains in passive margin salt basins: Insights from analogue modelling, Solid Earth, 10, 1283-1300, https://doi.org/10.5194/se-10-1283-2019.

32 https://www.auscope.org.au/agn

33 https://www.earthchem.org/

34 http://georoc.mpch-mainz.gwdg.de/georoc/

35 https://www.earthchem.org/communities/onegeochemistry/

Ge, Z., M. Warsitzka, M. Rosenau and R. L. Gawthorpe (2019). Progressive tilting of salt-bearing continental margins controls thin-skinned deformation, Geology, 47, 12, 1122-1126, https://doi.org/10.1130/G46485.1.

Hanson, B., K. L. Lehnert and J. Cutcher-Gershenfeld. (2015). Committing to Publishing Data in the Earth and Space Sciences. In Eos, 96, American Geophysical Union (AGU), https://doi.org/10.1029/2015eo022207.

Heidorn, P. B. (2008). Shedding Light on the Dark Data in the Long Tail of Science, In Library Trends, 57, 2, 280-299, Project Muse, https://doi.org/10.1353/lib.0.0036.

Michail, M., M. Rudolf, M. Rosenau, A. Riva, P. Gianolla and M. Coltorti (2021). Shape of plutons in crustal shear zones: A tectono-magmatic guide based on analogue models, J. Struct. Geol., 150, 104417, https://doi.orgM/10.1016/j.jsg.2021.104417.

Newman, J., R. T. Williams, N. Roberts, N. J. Phillips, A. D. Lusk, J. Ash, B. Tikoff (2021) StraboMicro: An application for sharing and contextualizing micro-scale geologic data. AGU Fall Meeting 2021, https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/997544.

Román-Berdiel, T. A. M. Casas, E. L. Pueyo, A. Peiro, R. Soto, A. Pohlenz, M. Warsitzka and M. Rosenau (2019). Ring shear test data of quartz sand and colored quartz sand used for analogue experiments in the Analogue modelling laboratory of the University of Zaragoza, Spain (EPOS TNA call 2017), GFZ Data Services, https://doi.org/10.5880/fidgeo.2019.025.

Rudolf, M., M. Rosenau and O. Oncken (2021). Ring Shear and Slide-Hold-Slide Test Measurements for Soda-Lime Glassbeads of 300-400µm diameter used at the Helmholtz Laboratory for Tectonic Modelling, Potsdam, Germany [Data set]. GFZ Data Services. https://doi.org/10.5880/GFZ.4.1.2021.002.

Stall, S., L. R. Yarmey, R. Boehm, et al. (2018). Advancing FAIR data in Earth, space, and environmental science. – Eos, Earth Space Sci. News, 99, https://doi.org/10.1029/2018EO109301.

Ulbricht, D., K. Elger, B. Radosavljevic and F. Ott (2020). Long-tail data curation in the times of the FAIR Principles and Enabling FAIR Data – challenges and best practices from GFZ Data Services, EGU General Assembly 2020, Online, 4-8 May 2020, EGU2020-16466, https://doi.org/10.5194/egusphere-egu2020-16466.

Ulbricht, D., K. Elger, T. Tesei, D. Trippanera (2017). The XML Metadata Editor of GFZ Data Services. Geophysical Research Abstracts, Vol. 19, EGU2017-12526-1, 2017, EGU General Assembly 2017. URL: https://meetingorganizer.copernicus.org/EGU2017/EGU2017-12526-1.pdf (last access 24 Nov 2021).

von der Linden, J., C. Kimblin, I. McKenna, S. Bagley, H.-C. Li, R. Houim, C. S. Kueny, A. Kuhl, D. Grote, M. Converse, C. E. J. Vossen, S. Stern, C. Cimarelli, C. and J. Sears (2021). Standing shock prevents propagation of sparks in supersonic explosive flows, Commun. Earth Environ., 2, 195, https://doi.org/10.1038/s43247-021-00263-y.

Wessels, R., G. ter Maat, E. del Bello, L. Cacciola, F. Corbi, G. Festa, F. Funiciello, G. Kaviris, O: Lange, J. Lauterjung, R. Pijnenburg, G. Puglisi, D. Reitano, C. Ronnevik, P. Scarlato and L. Spampinato (2022) Transnational access to facilities for Solid Earth Sciences: a service to promote multi-domain research in Europe, Annals of Geophys, 65, 2, https://doi.org/10.4401/ag-8768.

Wilkinson, M., M. Dumontier, I. Aalbersberg, I. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship, Sci Data. 3, 160018, https://doi.org/10.1038/sdata.2016.18

***CORRESPONDING AUTHOR: Kirsten ELGER**,
GFZ German Research Centre for Geosciences, Potsdam, Germany,
e-mail: kirsten.elger@gfz-potsdam.de