ANNALS OF GEOPHYSICS, 69, 2026; doi:10.4401/ag-9546
OPEN ACCESS

A structured metadata model for sharing data from rock deformation laboratory experiments

Stefano Aretusini*[,1], Manuela Sbarra[1], Massimo Cocco[1]

(1) Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy

Article history: received February 14, 2026; accepted July 10, 2026

Abstract

Structured metadata defined in a machine-readable format are the foundation of FAIR data and open research. To describe rock deformation laboratory experiments, standardized metadata formats are currently missing. This limits the annotation of metadata describing the experiments, as well as the data and metadata produced during the experiments. To address this issue, we propose a structured metadata model, a configuration for OpenBIS data platform, and a web interface to annotate the metadata of experiments and data files. We developed a structured metadata model based on our user experience on performing rock deformation experiments. Then, we fully configured OpenBIS data platform using a docker container stack. Finally, we modified an existing Streamlit application allowing users to annotate metadata according to our structure, upload and annotate data files automatically. This approach allows us to annotate data and metadata produced in rock deformation laboratory experiments in a systematic and structured way to foster the adoption of FAIR principles and open science across rock deformation laboratories and hopefully across geosciences laboratories in general.

Keywords: Metadata; Data; Mineral and rock physics; Open research; OpenBIS

1. Introduction

The ambition to make FAIR (Findable, Accessible, Interoperable and Re-usable) principles the basic standard for scientific research is nowadays a requirement to fulfil Open Science and the sharing of research data (Wilkinson et al., 2016). In this framework, FAIRness of research data is a necessary target for research infrastructures in all scientific domains and at global level, such as the European Open Science Cloud (EOSC) portal launched in November 2018. The European solid earth science community decided to address this challenge through the design, implementation, construction, and operation of the EPOS (European Plate Observing System) Research Infrastructure (Bailo et al., 2022, 2023; Cocco et al., 2022, 2025; Saleh Contell et al., 2022). EPOS is a Research Infrastructure (RI) committed to enable excellent science through the integration, accessibility, use and re-use of solid Earth science data, research products and services, as well as by promoting physical access to research facilities (Cocco et al., 2022; Cocco and Montone, 2022). The organizational model adopted to implement and operate the EPOS RI relies on EPOS ERIC (European Research Infrastructure Consortium) being the center for the coordination and management of geographically-distributed observing systems and laboratories, remotely accessible computational resources, and data repositories (Atakan et al., 2022; Saleh Contell et al., 2022). Sharing and access to high-quality data requires effective and sustainable management over the entire data lifecycle, which must be done by data providers belonging to the diverse scientific communities participating in the EPOS integration plan. To this task, EPOS established the Thematic Core Services (TCS) representing the community-specific integration layer participating to the EPOS governance framework (Cocco et al., 2022) that is needed to ensure the data and service provision (i.e., access to data, data products, services and software). There are ten Thematic Core Services in EPOS: the Multi-Scale Laboratories (MSL) is one of them, coordinating a network of European experimental laboratories bringing together the scientific fields of these subdomains: analogue modelling of geological processes, geochemistry, field-scale laboratories, rock and melt physics, magnetism and paleomagnetism, and microscopy and tomography (Elger et al., 2022). Most of the EPOS TCS integrate data and services from national, regional, and local observing systems (such as seismic, geodetic, and volcanic monitoring networks, or geomagnetic observatories) that include data centers for the management and curation of data and metadata. Few data repositories integrate data from experimental laboratories or local infrastructures that do not have data centers for the integration, management, and curation of research data. The latter generates data belonging to the so-called "long tail of science" (Heidorn, 2008; Stahlman and Kouper, 2026). The availability of rich metadata from experimental laboratories, requiring a standard metadata model that is currently missing, is the necessary pre-requisite to address service-driven data provision for the sharing of research data, going beyond the sharing of data associated with scientific publications. There are numerous services operating through the EPOS Platform to deliver very diverse research data which are described by more than 20 types of metadata (Bailo et al., 2022, 2023). The present article addresses the challenge of presenting a metadata structure for managing, storing, using, and sharing data from experimental rock-physics laboratories.

Structured metadata defined in a machine-readable format are the foundation of FAIR data management and open science since they enable the sharing of research data among scientists through webservices. In this framework, large data sharing platforms such as EPOS Platform[1] (Bailo et al., 2023) rely on webservices providing machine-readable structured metadata which are currently absent in the case of rock deformation experimental data. Examples of sound metadata models are in the seismological community, where metadata models allowed for building of data reduction code and pipelines exploiting the metadata objects (e.g. ObsPy (Beyreuther et al., 2010)) and building sound webservices for data access ever since the dawn of internet era (International Federation of Digital Seismograph Networks (FDSN) web services (DzIewonski, 1994)). On top of this, scientists now need data collection, reduction, modelling, and publication semi-automatic pipelines for a time effective and reproducible publishing of data, data quality control, etc. Finally, the high paced development of artificial intelligence requires machine-readable metadata for training machine learning models, neural networks, or large language models.

Research data produced in rock deformation laboratories are the foundations upon which many other geophysics, tectonics and structural studies are based. Rocks permeability data, for instance, obtained in the lab are used for models to interpret crustal strength (Townend and Zoback, 2000), hydrogeological and geochemical fluid migration processes (Brace, 1980; Wibberley et al., 2008). The P and S wave velocity obtained from lab testing are then used in tomography models to interpret the crustal changes of seismic velocity due to the lithologies occurring at depth and the changes related to fluid migration or poroelastic stress changes (Christensen and Mooney, 1995). The constitutive parameters governing the earthquake rupture process obtained from laboratory friction experiments are used as material properties in dynamic rupture models which are then used to interpret earthquake rupture propagation (Scholz, 1998) and the recorded radiated seismic waves (Bizzarri, 2011). Nevertheless, in most rock deformation laboratories worldwide, metadata defining the materials, machines, sample holders, people, and control steps involved in the experiments are either written on paper sheets and prone to be lost or written on excel spreadsheets or other non-machine-readable file formats (Tenopir et al., 2011). When this is not the case, researchers from these laboratories publish datasets which have descriptions and metadata that might be internally coherent in that dataset but are unstructured, often not machine-readable. This results in difficulties arising when researchers compare individual datasets in meta-analyses but also results in a cumbersome activity when one researcher needs to describe manually their data upon publication of the dataset in a public repository.

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

Developing robust metadata standards requires a coupled data architecture that bridges bottom-up community workflows with top-down federated infrastructures. At the domain scale, specialized geoscience laboratories communities have formalized data schemas tailored to their domain field; examples include StraboSpot (including *StraboMicro* and *StraboExperimental*) for microstructural and rock mechanics data (Walker et al., 2019), MagIC for paleomagnetic data (Tauxe et al., 2016), and EarthChem, GEOROC (Lehnert et al., 2000), and AusGeochem (Boone et al., 2023) for geochemical data. Concurrently, large-scale international infrastructures, such as European Union's EPOS Multi-Scale Laboratories (MSL) (Elger et al., 2022), Australia's AuScope, and North America's IEDA2, act as top-down federations. These overarching platforms provide the unified governance and technical frameworks necessary to scale up and harmonize the various research domain data into a globally interoperable, FAIR data ecosystem (Prent et al., 2023). In recent years, integrating the diverse data from the MSL TCS geoscience laboratory community into the EPOS platform was a challenging task. One of the potential causes of this was the lack of resources to develop sustainable data management in the experimental laboratories and the delays in implementing rich and structured metadata to be used for fostering and enabling interoperability among data centers. To address this issue and mitigate the effects of the lack of sustainable solutions for data interoperability, the MSL TCS decided to adopt as solution the link of data to publications through the provision of a Digital Object Identifier (DOI). Therefore, current metadata models for rock deformation experiments and laboratory experiments in the field of geosciences and geophysics in general are focused on the publication on scientific journals and DOI attribution (i.e., DOI minting) to the dataset for quick publication in data repositories (e.g., GFZ metadata editor[2]) at the time of publication. This is not the optimal solution for data interoperability, and it leads to a few problems: i) the aggregation of multiple types of data relative to multiple individual experiments and ii) the publication of sometimes poorly detailed metadata attached to .ZIP files containing unstructured data whose content is not discernible in machine-readable format. Other issues are lacking a shared ontology, absence of permanent identifiers attributed to metadata descriptors, etc. The absence of permanent identifiers (PIDs) is a shared problem between all laboratory facilities, which is partly being tackled by attributing PIDs (such as DOIs) to scientific instruments[3]. With the current TCS MSL efforts to promote TransNational Access (TNA), i.e. researchers mobility across laboratory facilities in EU (Wessels et al., 2022), by means of platforms such as the FAST[4] or the ILGE[5] portals, a systematic and structured method to manage and share data and metadata is furthermore needed. Previous work in EPOS TCS MSL community provided a geoscience laboratory community-based vocabulary[6], even if a comprehensive ontology (Noy and McGuinness, 2001) and metadata model built upon that vocabulary is currently lacking. In countertrend with this, in some rock deformation communities (European Society for Deformation Mechanisms, Rheology and Tectonics[7], Physical Properties of Earth Materials[8], or Collaborative Organization for Rock Deformation[9]) there have been recent efforts aimed at developing a consistent and structured metadata model leading to the production of tools such as StraboExperimental[10] (Walker et al., 2019) and workflows such as LAPS (Mok and Pec, 2021). Moreover, geochemistry community with GEOROC database[11] (Lehnert et al., 2000) and paleomagnetism community with MagIC database[12] (Tauxe et al., 2016) proceeded in the last years in a similar direction with remarkable success.

A successful example of data and metadata sharing is in seismology the FDSN[13] (DzIewonski, 1994) that since several decades federated and standardized all producers of seismogram data, allowing for establishing a common

2 https://dataservices.gfz-potsdam.de/elmo/.

3 https://docs.pidinst.org/en/latest/white-paper/index.html.

4 https://fast.geo.uu.nl/app/login.

5 https://ilge.ct.ingv.it/index.html.

6 https://github.com/UtrechtUniversity/msl_vocabularies.

7 https://drt-society.org/.

8 https://connect.agu.org/mrp/weblinks/ppem.

9 https://mpec.scripts.mit.edu/peclab/cord/.

10 https://www.strabospot.org.

11 https://georoc.eu/georoc/new-start.asp.

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

13 https://www.fdsn.org/about/.

language, metadata model and data reduction method (e.g., ObsPy (Beyreuther et al., 2010)). Indeed, seismology might be an example of "head of science" as monitoring networks produce centralized data that since their origins were forced to be made uniform and standardized. On the other end the "tail of science" is represented by individual laboratories in which data are more heterogeneous, user dependent and condition-specific and more difficult to standardize. Nevertheless, this difficulty should not be a limiting factor in proceeding in the standardization direction. In other scientific and engineering communities whose main specialty is laboratory experiments, the focus on developing a metadata model that can describe the experiment already occurred, fostered by community building. In biology or material science communities the advantage in data standardization provided by usage of Electronic Laboratory Notebooks (ELNs) and Laboratory Information Management Systems (LIMSs) pushed for the development of communities and shared metadata models (i.e., for biology (Bauch et al., 2011), for material sciences (Siemer et al., 2022)). Similar trends were documented for other engineering related disciplines[14]. Moreover, we would like to highlight how the "rock and melt physics" and the material science community are similar in terms of ontology, instruments used and experimental procedures. Hence, we think that the rock and melt physics community should drive this push for publishing structured metadata and data in EPOS MSL TCS.

In this paper, we present the research activities supported by the Joint Research Unit EPOS-Italy aimed at implementing a structured metadata model for sharing data stored at the rocks and earthquakes mechanics laboratory in the HPHT laboratories of the Istituto Nazionale di Geofisica e Vulcanologia (INGV, Rome). The motivation behind the initiative is twofold: first, to create an infrastructure to manage and archive data from the rock and earthquake mechanics laboratory at INGV, and second, to develop services that promotes findability and interoperability. Our effort aims at providing individual rock deformation laboratories with tools to implement a data management infrastructure based on a metadata model that can allow for FAIR data management at a very small scale while at the same time promote synergy with the broader data management scale of EPOS MSL and EPOS Platform. We included the existing MSL vocabulary[15] into a metadata and data model implemented in Openbis, an open source highly flexible data management system developed at ETH Zurich (Barillari et al., 2016; Bauch et al., 2011). We configured Openbis as a dockerized multi-container application. We developed an existing Streamlit app (Rejiba et al., 2024) and used it as a web page front end to annotate metadata of the experiments and extract additional metadata from data files. We present here one deformation experiment carried out on a rotary shear machine SHIVA (Di Toro et al., 2010) as an example of the implemented metadata structure, exported and published in Zenodo (Di Toro et al., 2025). We highlight the importance of a standardized description of experimental laboratory data to implement FAIR data principles in our rock deformation community.

2. Methods

2.1 Proposed IT architecture

We implemented an IT architecture to store data and annotate the associated metadata for the rock and earthquake mechanics laboratory of the INGV HPHT laboratories in Rome. We present our local data management infrastructure in the framework of the EPOS MSL IT architecture (Fig. 1). At its current state, our Openbis data management infrastructure is partly like Lab Type C (Fig. 1a) as Openbis has a native export interface to Zenodo. Interface of our Openbis data management infrastructure (step 1, level C) to a Data Repository may occur by aggregation of data and metadata (step 2). Alternatively, the interface of our Openbis data management infrastructure directly to the EPOS Platform (step 5) may occur by means of a dedicated web service. The further implementation of these issues is currently under discussion, and it will be briefly addressed in this paper.

14 https://www.rd-alliance.org/disciplines/engineering-and-technology/.

15 https://github.com/UtrechtUniversity/msl_vocabularies/tree/main.

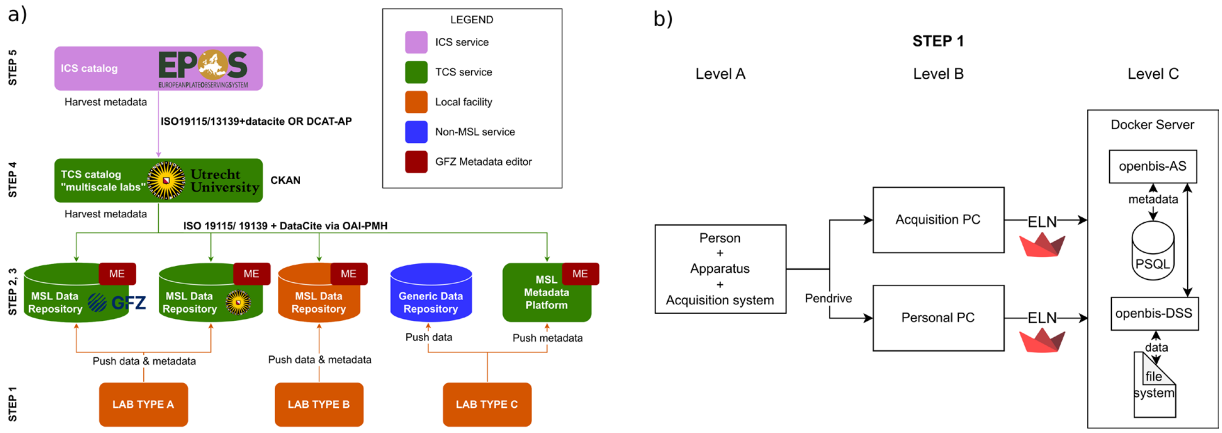

Figure 1. Our IT architecture in the framework of EPOS MSL. (a) EPOS MSL IT architecture diagram[16]. Step 1: individual laboratories producing data; step 2: metadata and data are annotated (e.g. in the GFZ metadata editor); step 3: metadata and data are published in Data Repositories with attribution of DOI (i.e., DOI minting); step 4: the Thematic Core Services (TCS) repository harvests metadata and data from the Data Repositories; step 5: the EPOS Platform harvests data and metadata from the TCS repository. (b) A detail within step 1, with our application, highlighting the data and metadata path from the source, where data is produced by people using apparatuses and acquisition systems (level A), to where data is temporarily stored, in a PC attached to an apparatus or personal (level B), to our Openbis data management infrastructure (level C). Interface B to C can occur with either the Electronic Lab Notebook natively installed with Openbis or with the modified Streamlit app, tailored to our metadata model.

2.2 Openbis installation and metadata model

Openbis (Barillari et al., 2016; Bauch et al., 2011) architecture included a stack of docker containers: i) PostgreSQL database, ii) Openbis including both the Application Server (AS) and Data Store Server (DSS) services, iii) Lightweight Directory Access Protocol (LDAP) authentication system with iv) its LDAP manager web interface. A separate stack included the Streamlit app. Both stacks were installed in a Docker Server instance hosted by INGV IT services.

The metadata model in Openbis is composed of hierarchic containers, which are from the largest to the smallest: Instance, Space, Project, Collection/Experiment, and Sample/Object. Samples can be in a parent-child relationship with each other's. In this framework, Samples and Object, Collection and Experiment are synonyms so we will use Object and Experiments in the next sections of the text. Property and Property Types are used to describe each of the items above. Our custom metadata model is composed of Experiment and Object Types and their relative Properties (belonging or not to a Property Type). Data Set Types are used to describe the actual, file-based, data by means of their Properties. Data Sets can be nested into an Experiment or an Object and can contain one or more files and their relative Properties. Properties can be vocabularies of a specific Vocabulary Type. In the following sections the Experiment, Object, Data Set or Vocabulary types will be in the uppercase.

We developed our metadata model considering the needs of the rock deformation community of several Object Types to describe experiments. The DEFAULT_EXPERIMENT Experiment Type represents each experiment here. The metadata model included Object Types describing the materials, specimens, instruments, people, and experimental steps involved in the experiments (see Section 3.1.1). We developed Data Set Types to describe column-based data files, images according to metadata extracted with Exiftool[17], and data acquisition metadata (see Section 3.1.2). To describe in a standardized way some Object Type Properties, we included the MSL vocabulary (v.3) that we divided into seventeen sub-vocabularies, and more vocabularies to describe the specific types of: physical quantities, measurement units, artificial materials, fluids, and rocks (see Section 3.1.3).

16 https://www.epos-eu.org/tcs/multi-scale-laboratories/data-services.

17 https://exiftool.org/.

The metadata and data model (all Types described above) were imported in Openbis via the admin UI after the initial installation. We acknowledge the availability of a metadata structure schema importer (Rejiba, Korte-Kerzel et al., 2025) and other tools for metadata model versioning and conflict resolution[18].

The Streamlit app was forked and adapted from an existing repository developed by material scientists (Rejiba et al., 2024), and consisted in several pages: i) user authentication with a personal access token (that is obtained from the Openbis ELN); ii) experiment metadata annotation by editing and choosing Object Types; iii) registration and upload of data to Openbis DEFAULT_EXPERIMENT, with automatic metadata annotation into Data Sets performed with metadata parsers. Our work here focuses on implementing in Data Set metadata parsers for the automatic extraction of metadata in our custom Data Set Types and relative Properties.

Starting from the SHIVA apparatus laboratory experiments paper sheets, we translated one SHIVA experiment (s0050) into our custom metadata structure and exported it to Zenodo directly from Openbis (see Data availability).

3. Results

3.1 Structured Metadata and Data model

Natively, Openbis metadata model is represented by hierarchical types, from the highest to the lowest level, where each upper level is a container of the lower. Types are named: Instance, Space, Project, Collection (or Experiment), and Sample (or Object). A Data Set can be contained in Experiments or Objects.

In this framework, Object Types define the building blocks of the metadata structure and can be set in parent-child relationship. Each Object implements Properties among which are Vocabularies. To define the metadata model, we included abstract categories as controlled vocabularies, whereas usable items were included as Objects. We show and describe here the list and meaning of all Objects composing the metadata schema, and the entity relationship diagrams with Objects and their Properties (Figs. 2, 3, 4, and 5; Supplementary File S1).

3.1.1 Object types and their relationships

The Object Types composing the metadata schema can be grouped logically into these categories: basic materials, specimens (solid or powdered), instruments, people, experimental steps (basic or specialized) and derived measurements (Figs. 2, 3, 4, and 5; Supplementary File S1).

Basic material types included several types of samples, either bought or sampled from the field (Fig. 2):
- Samples of artificial materials (ARTIFICIAL), e.g., a steel plate, PTFE plate, PTFE tube.
- Core samples (CORE), e.g. a IODP core (from Methods in (Fisher et al., 2005)).
- Fluid samples (FLUID), e.g., a fluid sampled from the field or a bottle of commercial fluids.
- Rock hand samples (ROCK), e.g., a rock sampled from the field, or a slab purchased from a quarry.
- Sediment samples (SEDIMENT). e.g., a sediment sampled from the field, a sample of natural granular materials or soil.

Each Basic material can be set as parent of **powdered material** types that included:
- Powder samples (POWDER), e.g., any material above after being ground or powdered and sieved.
- A utility object (POWDER_PROPORTION) linking with a 1:1 relationship a powder sample with its percent weight. One or more POWDER_PROPORTION can be set as parent of a mixture of powder samples (POWDER_MIXTURE) to define how each powder is mixed in terms of weight percent into the powder mixture.

18 https://github.com/BAMresearch/bam-masterdata/tree/main.

Specimens included any basic (except FLUID) or powder materials that can be transformed into a 3D geometric shape.
Solid specimen types (Fig. 2) can be set as children of ROCK, CORE or ARTIFICIAL are:
- hollow cylinder specimens (SOLID_SAMPLE_HOLLOWCYLINDER),
- cylinder specimens (SOLID_SAMPLE_CYLINDER),
- block specimens (SOLID_SAMPLE_BLOCK).

Powdered specimen types (Fig. 2) can be set as children of POWDER, POWDER_MIXTURE are:
- generic shapeless powder specimens (POWDERED_SAMPLE),
- hollow cylinder powder specimens (POWDERED_SAMPLE_HOLLOWCYLINDER),
- cylinder powder specimens (POWDERED_SAMPLE_CYLINDER),
- and block powder specimens (POWDERED_SAMPLE_BLOCK).

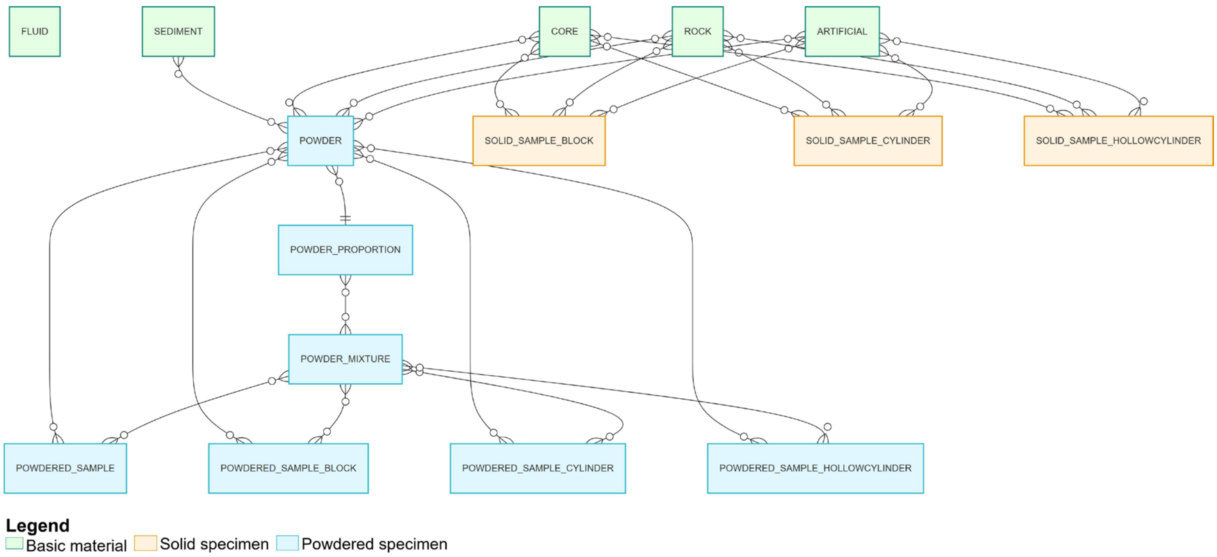

Figure 2. Object types represent basic materials (light green), solid specimens (orange) and powdered materials and specimens (blue). The possible parent-child relationships are indicated. This part of the metadata schema is used to represent: i) actual items sampled from the field or purchased from a company (basic materials), ii) solid specimens with their geometrical shapes derived from (i.e., child of) the basic materials iii) powders derived from (i.e., child of) the basic materials, iv) specimens with their geometrical shapes derived from (i.e., child of) powders.

Instrument types included (Fig. 3):
- Machinery not producing data (EQUIPMENT), e.g., balance, drill, heating plate.
- Machinery producing data (APPARATUS), e.g., a triaxial, biaxial, or rotary apparatus.

APPARATUS can be set as parent of:
- Machinery applying a physical change to a specimen (ACTUATOR) e.g., a piston, valve, pump.
- A collection of parts defining the specimen holder (ASSEMBLY), e.g., a Hooke's triaxial cell. ASSEMBLY can be set up as parent of the individual parts (COMPONENT) constituting the specimen holder e.g., a rubber jacket, or a graphite oven. COMPONENT types can be set as child of ROCK or ARTIFICIAL.
- Instruments recording a physical change (SENSOR) e.g., a linear variable differential transducer, pressure transducer, piezoresistive transducer.

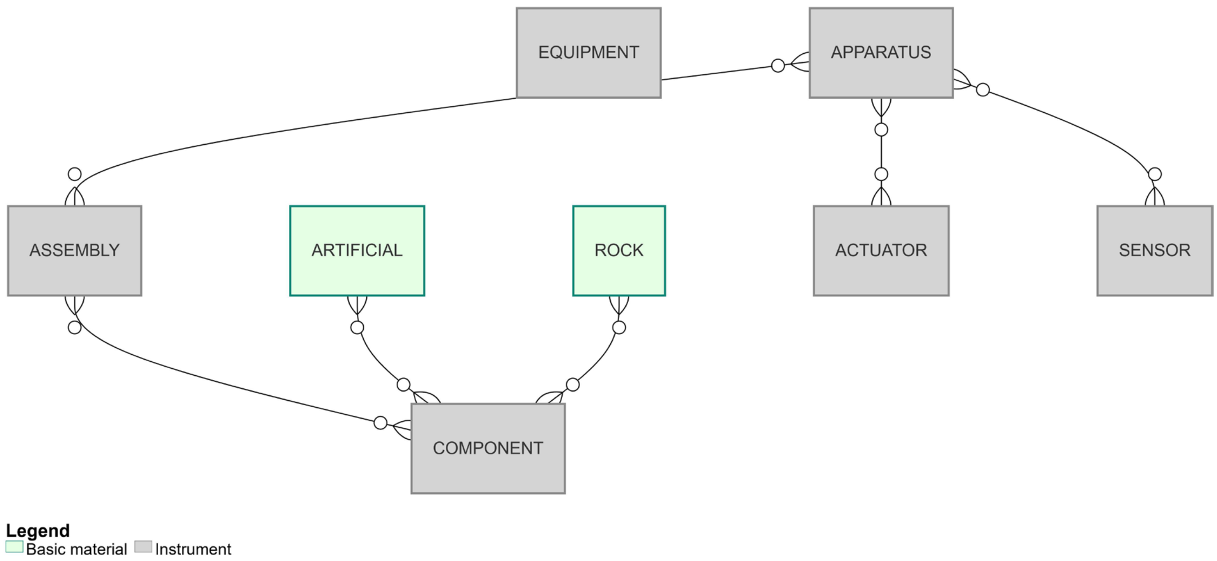

Figure 3. Object types representing instruments (grey) or basic materials (light green). The possible parent-child relationships are indicated. This part of the metadata schema is used to represent the machinery (or hardware) used in the experiments. Experimental or sample assemblies, actuators, sensors are attributed to (i.e., are child of) one or more experimental apparatuses. Components are the items that comprise (i.e., are child of) an experimental or sample assembly, and can be made of (i.e., are child of) artificial or rock basic materials.

People type included the actual people executing or assisting to an experiment (PERSON) (Fig. 4).

Experimental step types are implemented as sequences of basic and specialized steps in a hierarchic parent-child structure. The first step of the structure is child of other objects described above, i.e., PERSON, SENSOR, APPARATUS, EQUIPMENT, ASSEMBLY, ACTUATOR, FLUID, any powdered or solid specimen type (Fig. 4). The first step is parent of a sequence of basic or specialized steps providing a chronological order of execution (Fig. 5).

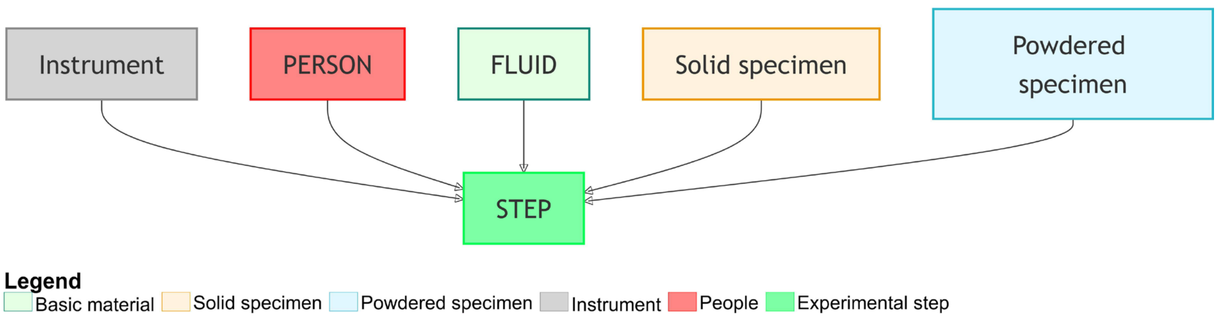

Figure 4. Object types that can be parents of the first basic experimental step (green): instruments (grey), PERSON (red), FLUID (light green), solid specimens (orange), and powdered specimens (blue). In this manner, all the items and people involved in the experiment are linked to the first basic experimental step and therefore inherited by the entire sequence of experimental steps.

Basic steps (STEP) are generic experimental steps (or stages), e.g., heating stage, running-in step.

Specialized steps can be children of the ACTUATOR applying/controlling them and of a basic STEP, can be named after a mathematical function (STEP_CONSTANT, STEP_SINUSOIDAL, STEP_TRIANGLE, STEP_REGYOFFE, STEP_BOXCAR, STEP_RAMP, STEP_STEPS) or represent a pause (STEP_HOLD) e.g., a constant with value 0. Specialized steps can be generic across several Apparatuses with an Actuator piston (STEP_TOUCHPOINT) e.g., generic touchpoint procedure with force increase due to contact of piston on specimen, or can be tailored for a specific Apparatus such as SHIVA:

velocity control step (STEP_TRAPZ_TIMELIMITED), position control step (STEP_TRAPZ_SLIPLIMITED), or torque control sequence of steps (STEP_STEPS_TORQUE).

Derived measurement types are extracted from the experiment data and can be the child of any basic or specialized steps and the PERSON deriving the measurement. The derived measurement can be a point (POINT_MEASUREMENT) e.g., a friction point measurement.

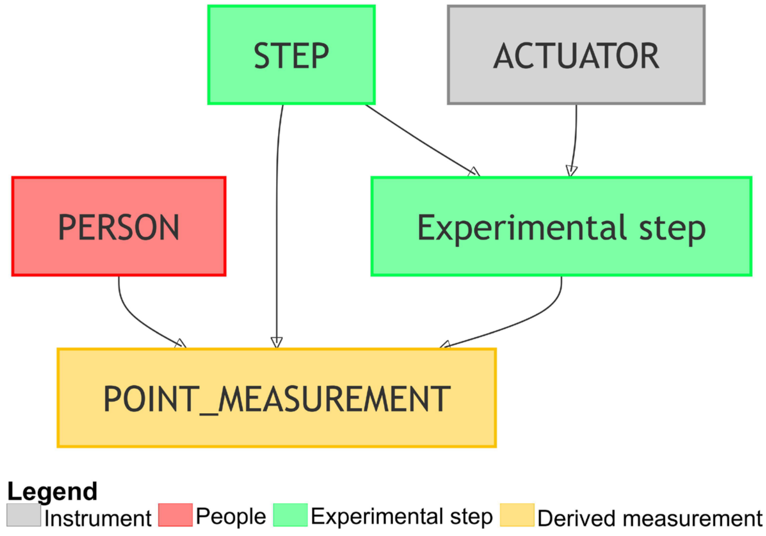

Figure 5. Object types representing basic steps and their specialized step (green) or derived measurement (yellow) children. The possible parent-child relationships are indicated. In this manner, each specialized experimental step is linked to (i.e., is child of) the basic step indicating one stage of the experiment and the actuator used to perform the control action. At the same time, each derived measurement (e.g., a permeability or friction value) is linked to (i.e., is child of) the experimental step (basic or specialized) and to the person performing the measurement.

3.1.2 Data Set types

The metadata of actual Data Sets is represented by Data Set Types and their Properties. We developed the following specialized Data Set Types (Supplementary File S2):

- Implementing either 16, 64 or 128 header names mapped the relative measuring SENSOR (16HEADERS, 64HEADERS, 128HEADERS).
- Implementing exiftool metadata (IMAGE).
- Implementing apparatus- and software-specific metadata, employing a naming convention as <apparatus/equipment name>_<software name (when present or known)>_ACQUISITION_METADATA (SHIVAMOTOR_ACQUISITION_METADATA, SENTEA_PEAKVIEWER_ACQUISITION_METADATA, RICHTER_INSITE_ACQUISITION_METADATA).

3.1.3 Vocabulary types

We included the following vocabularies in the metadata model (Supplementary File S3). We included the MSL vocabulary[19] that was divided into 17 sub-vocabularies: MSL_EQUIPMENT, MSL_MEASUREDPROPERTY, MSL_MINERAL, MSL_ARTIFICIAL, MSL_ROCK, MSL_SEDIMENT, MSL_POREFLUID, MSL_INFERRED, MSL_SETTING, MSL_MODEL, MSL_SOFTWARE, MSL_ANALYSIS, MSL_FACILITY, MSL_FACILITYTYPE, MSL_GEOLOGICALAGE,

19 https://github.com/UtrechtUniversity/msl_vocabularies/tree/main, version 1.3.

MSL_MICROSTRUCTURE, MSL_ENVIRONMENTCONTROL. We included also more widely used vocabularies to describe other Properties such as the specific types of physical quantity (QUDT_QUANTITYKIND)[20], measurement unit (QUDT_UNIT)[21], artificial material (SWEET_MATR_IND)[22], fluid physical state (SWEET_STATE_PHYS)[23], and rock (BGS_ROCKNAME)[24]. Each vocabulary had Code and Description Properties. Description included the URI of the implemented vocabulary. Each term in each vocabulary had Code, Label and Description Properties. The Description included the URI of each term. Terms from the implemented vocabularies are used as Properties of some of the Object Types. We use in our metadata model: MSL_SEDIMENT in SEDIMENT Object, MSL_ROCK and BGS_ROCKNAME in ROCK Object, MSL_EQUIPMENT in APPARATUS, EQUIPMENT, ASSEMBLY, ACTUATOR, COMPONENT, and SENSOR Objects, QUDT_UNIT and QUDT_QUANTITYKIND in SENSOR Object, MSL_ARTIFICIAL and SWEET_MATR_IND in ARTIFICIAL Object, MSL_POREFLUID and SWEET_STATE_PHYS in FLUID Object, MSL_MEASUREDPROPERTY in POINT_MEASUREMENT Object. QUDT_UNIT was also used in the specialized steps Object Types.

3.2 Structured Logic in Openbis (Openbis masterdata)

We describe here how we envision the structure of the Openbis data management infrastructure (Supplementary File S4) and how it relates to EPOS-MSL. Each laboratory (Step 1 in Fig. 1a) implements its Openbis. In Openbis, in the Lab Notebook Instance, each Space is specific of one apparatus performing experiments and collecting data. For each Space, one or more Projects collect the laboratory experiments. Each laboratory experiment is represented by a Collection/Experiment (DEFAULT_EXPERIMENT), that contains all the experimental steps helping to define the sequence of actions taken during the experiment. Each DEFAULT EXPERIMENT contains one or more Data Sets that explain the actual data acquired during the experiment (each Data Set can be one single file or a folder). Together with the experiments, also collection of specimens can be created in each project, to include some specimens that are not shared across all apparatuses Spaces. In Openbis, in the Inventory Instance, Spaces and Projects contain the Collections of Objects that collect the digital twins of the physical Objects that help define the laboratory experiment (i.e., the parents of the first step of each experiment).

The access of apparatus-specific Spaces in the Lab Notebook is determined by the user authentication structure with a granularity defined at the Project level. On the other hand, all Collections of Objects in the Openbis Inventory are accessible, editable, and therefore shared across apparatuses and people annotating their experiments.

3.2.1 Mandatory and optional elements to describe a laboratory experiment

In Section 3.1 we provided a full description of the possible Object Types that can be implemented. Here we define the mandatory set of Object Types that we deem necessary to define a laboratory experiment metadata. The laboratory experiment coincides with a DEFAULT_EXPERIMENT Collection/Experiment which Name Property is the internal identifier of the experiment employed traditionally in the lab (e.g. s0020 with “s” referring to SHIVA apparatus). Contained in the DEFAULT_EXPERIMENT, a minimum of one STEP Object is required with optional specialized step Objects as children. Alternatively, one specialized step can be implemented instead of the STEP Object (Fig. 6). Each specialized step Object has a compulsory ACTUATOR Object parent which has a compulsory APPARATUS Object parent. The first STEP or specialized step Object within the DEFAULT_EXPERIMENT requires among its parents at least one PERSON, one APPARATUS, one among solid or powdered specimen Objects. Solid specimen Objects require as parent one CORE, ROCK or ARTIFICIAL Object. Powdered specimen Objects require as parent one POWDER or POWDER_MIXTURE. POWDER requires as a parent one SEDIMENT, CORE, ROCK or ARTIFICIAL Object. POWDER_MIXTURE requires as parent a POWDER_PROPORTION Object. In general, FLUID, SENSOR, ASSEMBLY, COMPONENT, EQUIPMENT, and derived measurement Objects are Optional.

---

20 https://www.qudt.org/doc/DOC_VOCAB-QUANTITY-KINDS.html, version 3.1.10.

21 https://www.qudt.org/doc/DOC_VOCAB-UNITS.html, version 3.1.10.

22 http://sweetontology.net/matrIndustrial/, version 3.6.0.

23 http://sweetontology.net/statePhysical/, version 3.6.0.

24 https://vocab.bgs.ac.uk/v/vocab/rthmtrlclss:RockName, version 2.5.

3.3 Implementation of metadata model on existing data

The Streamlit app guides the user to implement the mandatory metadata structure thanks to web forms tailored on the metadata model and schema (Aretusini and Sbarra, 2026). This approach was chosen because such a structure to be filled from Openbis or Openbis ELN would have resulted in a cumbersome and time-consuming activity including the use of Excel templates. Starting from a .PDF scan describing the experiment (Fig. 6a), the user employs the Streamlit app and obtains a .JSON file (Fig. 6b). The .JSON file contains a series of newly defined steps objects that are in parent-children relationships defining the order of execution during the experiments, with the first step having as parents all other objects chosen from the user to define the experiment (people, materials, etc.), from the Inventory instance (Fig. 6b). This is then parsed into OpenBIS by the Streamlit app via pyBIS. The implementation of metadata can be done on-the-fly even during the experiment, including only the mandatory information (specimens, apparatus, people, and sequence of steps) and can later be enriched and/or corrected after the experiment.

The Data Set Types properties are annotated automatically upon file upload from the Streamlit app, by means of metadata parsers. This is also hastening the otherwise manual filling of Data Set properties. Metadata parsers must be defined prior to deploying the Streamlit app depending on the expected data produced in the laboratory (Fig. 6c). The result is a .JSON file describing all files metadata (Fig. 6d). This is then parsed into OpenBIS by the Streamlit app via pyBIS. The procedure of parsing metadata from files is normally done automatically and asynchronously from the metadata annotation as file formats and relevant metadata are not changing as frequently as the experiment metadata.

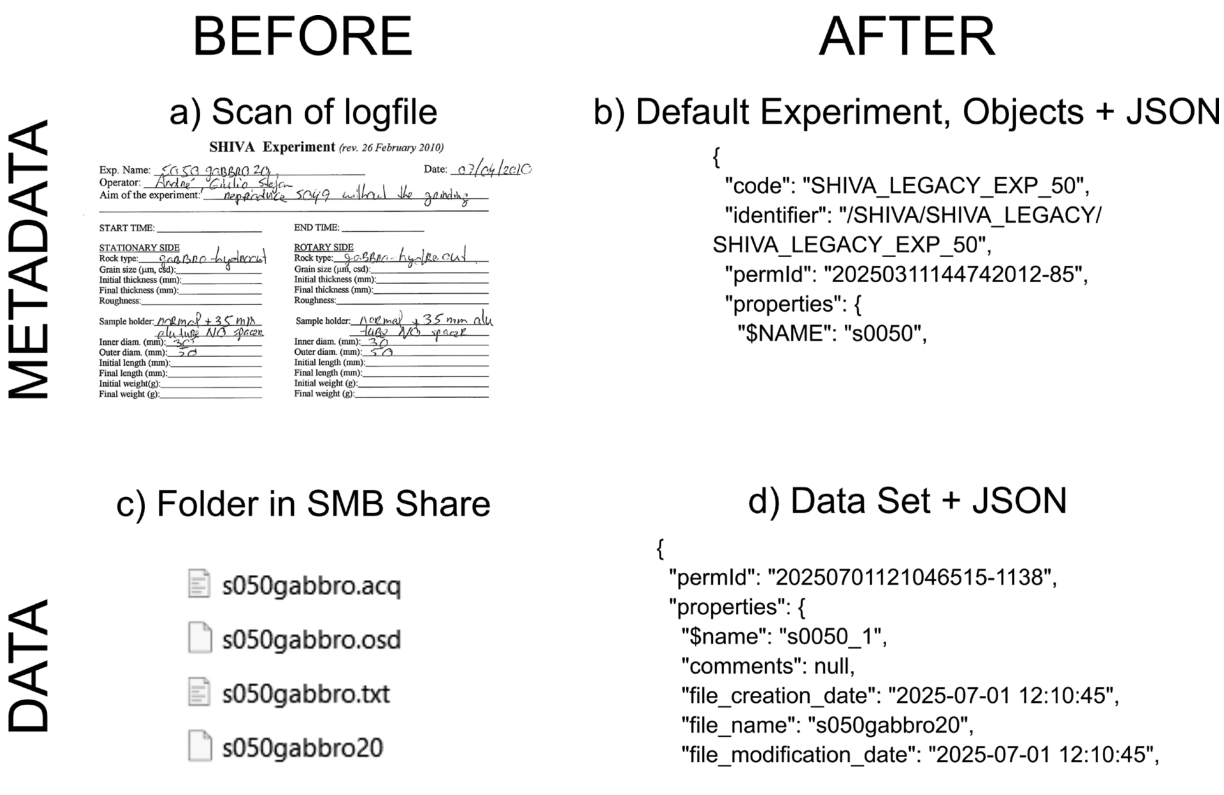

Figure 6. Implementation example for Shiva experiment s0050. Metadata: (a) before, (b) after and data (c) before and (d) after the application of our metadata model. The JSON are snippets of the entire JSON file in (Di Toro et al., 2025).

4. Discussion

Our structured metadata model tries to be as comprehensive and detailed as possible by capturing all the physical and abstract objects involved in a rock deformation experiment. However, being so specific and extensive is also complex. This complexity does not allow for direct metadata annotation from the native Openbis ELN, that would require manually filling cumbersome excel templates. Because of this, we customized the Streamlit app. In this way, most of the metadata and data complexity is handled by the Streamlit app Python code, which implements Pybis methods in the backend. A clear advantage of our metadata model can be seen comparing the example dataset s0050 (Fig. 6). Before, experiment metadata was in a .PDF, difficult to read by machines and not easy to read even by humans (due for example to complex calligraphy or lack of domain knowledge and jargon). After the application of the metadata model data is structured and machine readable. With respect to data files, stored in a SMB share, they did not contain metadata and were inaccessible, whereas after the metadata parsing, based on our metadata model, they are fully structured, machine readable and accessible via openBIS. This enables future implementation of web services for the interrogation of data via API, a necessary step to achieve integration with the EPOS Platform.

Albeit comprehensive, our structured metadata model surely lacks some objects that can be implemented when this model is expanded and generalized to other types of laboratories other than the rock deformation laboratory or for compatibility, for example, with LAPS or StraboExperimental (Section 1). This is entirely possible thanks to the high flexibility of Openbis. Moreover, the metadata parsers in the Streamlit app can be extended to many more Data Set Type metadata models or, for each Data Set Type, for more types of files for automatic metadata annotations by extending the python code that is in a modular form.

An analysis of Material Sciences metadata schema and workflows (Rejiba, Lee, et al., 2025) shows that our schema is fully compatible with those proposed by their community, after some adaptations are implemented. In their case, specimen Object Types (called “samples”) are children of the experimental step used for preparing or testing them, the experimental steps are children of the instrument involved and/or the experimental protocol which is child of the consumables used. Data Sets are then children of the sample. The focus on having a reproducible hierarchy representing sample (or specimen) history is because in material sciences, specimens are often synthesized and the relationship between multiple synthesis processes and final material properties needs to be clear and annotated in the metadata structure. On the other hand, in the geosciences laboratories, samples are collected from the field and are then transformed to obtain specimens for our testing machines. One way to make the two structures compatible would consist in adding a new customized specimen Object Type (e.g., such as a deformed specimen) as child of the last experimental step and of the specimen that is parent of the first experimental step so that the parent-child relationships in the specimen metadata tree are consistent.

The need for permanent identifiers for objects could also be an avenue for developing our data model and Streamlit app that is warranted by the flexibility of Openbis. Available examples could be the implementation of external identifiers as Properties of our Openbis Sample Types or as a dedicate Object Type binding existing Objects to external identifiers. Some of the likely candidates for implementations are: i) International Generic Sample Number (IGSN) identifiers for any basic material or specimen object (Klump et al., 2021); ii) Open Researcher Contributor ID (ORCID) identifiers for people and users (Haak et al., 2012), preferable to Web of Science (WOS), Scopus, or other institutional identifiers; iii) Research Organization Registry (ROR) identifiers for institutions or laboratories[25]; iv) Persistent Identification of Instruments (PIDINST) identifiers for instruments or apparatuses (Section 1). For some Object Types, also EPOS MSL identifiers may be available and can similarly be implemented in perspective (i.e., laboratories, facilities, equipment, sensors, software, etc.).

In Openbis, there are currently only two existing interfaces for the export of data and metadata: one with ETHZ Research Collection and another with Zenodo. Therefore, our future work will aim at developing API interfaces to a Data Repository (step 2-3 in Fig. 1) that is harvested by the TCS Data Repository, as well as to the EPOS Platform. In the first case, we would have to use the Data Repository to mint one DOI for each experiment similarly to what we did for the metadata and data of experiment s0050 but in an automatic way. In the second case, we would have to develop APIs to visualize experiments' metadata such as Maps (i.e., sample provenance) or Tables (i.e., all metadata) and experiments data as Graphs (i.e., plotting experimental variables against time).

Despite necessary, these developments raise a new challenge: the sustainability of research data management in an experimental laboratory including the interactions with users to ensure access, use, and re-use of data. This sustainability is not limited to the technical aspects mentioned above, it also includes the human resources and skills to manage research data during the entire lifecycle. A solution to this issue might be to connect the Openbis data management infrastructure to an institutional data center in charge of harmonizing, facilitating, and fostering the access and use of research data as well as addressing the users' requests. This solution also represents

25 https://ror.org/.

a sustainable way to optimize resources (human, technological and skills) at institutional level, to foster the adoption of shared institutional data policies and standardized web services, which in turn means supporting interoperability. This implies that the commitment of an experimental laboratory is dedicated to collecting, curate, and store high-quality research data and metadata, ensuring access, findability, and interoperability. The architecture presented here can be fully exported to other laboratories in the MSL community. Everyone in the community is welcomed to amend or discuss the metadata model leading to its consolidation and future development. This architecture is at the laboratory level, hierarchically below the TCS, and treats data with a greater degree of granularity with respect to the EPOS TCS data infrastructure. The aggregation of multiple data sets from multi-scale laboratories would require a research community coordination, which should be conducted by the EPOS MSL TCS. The architecture based on the structured metadata model built at the rock deformation and earthquake mechanics group of the INGV HPHT laboratory represents a pilot case to foster FAIR data management in small laboratories and therefore extract data from the "long tail of science".

5. Conclusions

We have presented a structured metadata model to foster the FAIR management of research data generated by the rock deformation and earthquake mechanics group at INGV Rome. This endeavor starts with recognizing the importance for researchers working in the experimental laboratory to make the data used for their scientific research and investigations Findable, Accessible, Interoperable, and Reusable (FAIR). To this task we implemented a data management infrastructure to collect, curate, store and manage laboratory data. This architecture ensures access and findability of laboratory data as well as going in the direction of enabling interoperability. The work presented in this paper focuses on the structured metadata model and the IT architecture of the data management infrastructure. Our experience highlights the importance of identifying further support to manage the data in the subsequent stages of the data lifecycle, when interfacing our data management infrastructure with an Institutional Data Center or a thematic community data center. This is because the management of users' interactions addressing their requests is beyond the capacities an individual experimental laboratory. This solution represents a sustainable approach to data management across the entire lifecycle. In this framework, our expectation is to contribute to FAIR data management and trigger a broader discussion in the rock deformation and in the geosciences laboratory community on the importance of developing a consistent and structured metadata model, built upon the one presented here, to help fulfill FAIR data best practices. We encourage the entire community to integrate, rediscuss, and modify our work with the goal of sharing geoscience laboratory data even independently from the research publication process.

Data availability statement. All metadata and data relative to SHIVA experiment s0050 are available in Zenodo (Di Toro et al., 2025). All files representing the metadata structure are available in Zenodo (Aretusini and Sbarra, 2026) and can readily be imported from Openbis admin UI for implementation. A high-resolution diagram of the entire metadata model is available as supplementary file S5.

Acknowledgements. We acknowledge JRU EPOS-Italia 2022 and 2023.

References

Aretusini, S. and M. Sbarra (2026). hpht-rockearthquakemechanics/openbis-rm-masterdata: Metadata model release (Version 1.0), Zenodo, doi:10.5281/ZENODO.20767223, last accessed 19 July 2026.

Atakan, K., M. Cocco, B. Orlecka-Sikora, R. Pijnenburg et al. (2022). National EPOS initiatives and participation to the EPOS integration plan, Ann. Geophys, 65, 2, DM211, doi:10.4401/ag-8758.

Bailo, D., K. G. Jeffery, K. Atakan, L. Trani et al. (2022). Data integration and FAIR data management in Solid Earth Science, Ann. Geophys., 65, 2, DM210, doi:10.4401/ag-8742.

Bailo, D., R. Paciello, J. Michalek, M. Cocco et al. (2023). The EPOS multi-disciplinary Data Portal for integrated access to solid Earth science datasets, Sci. Data, 10, 1, 784, doi:10.1038/s41597-023-02697-9.

Barillari, C., D. S. M. Ottoz, J. M. Fuentes-Serna, C. Ramakrishnan et al. (2016). openBIS ELN-LIMS: an open-source database for academic laboratories, Bioinformatics, 32, 4, 638-640, doi:10.1093/bioinformatics/btv606.

Bauch, A., I. Adamczyk, P. Buczek, F.-J. Elmer et al. (2011). openBIS: a flexible framework for managing and analyzing complex data in biology research. BMC Bioinformatics, 12, 1, 468, doi:10.1186/1471-2105-12-468.

Beyreuther, M., R. Barsch, L. Krischer, T. Megies et al. (2010). ObsPy: A Python Toolbox for Seismology, Seism. Res. Lett., 81, 3, 530-533, doi:10.1785/gssrl.81.3.530.

Bizzarri, A. (2011). On the deterministic description of earthquakes, Rev. Geophys., 49, 3, RG3002, doi:10.1029/2011RG000356.

Boone, S. C., F. Kohlmann, W. Noble, M. Theile et al. (2023). A geospatial platform for the tectonic interpretation of low-temperature thermochronology Big Data, Sci. Rep., 13, 1, 8581, doi:10.1038/s41598-023-35776-3.

Brace, W. F. (1980). Permeability of crystalline and argillaceous rocks, Int. J. Rock Mechan. Mining Sci. Geomech. Abs., 17, 5, 241-251, doi:10.1016/0148-9062(80)90807-4.

Christensen, N. I. and W. D. Mooney (1995). Seismic velocity structure and composition of the continental crust: A global view, J. Geophys. Res.: Solid Earth, 100, B6, 9761-9788, doi:10.1029/95JB00259.

Cocco, M. and P. Montone (2022). Preface to the Special Issue: EPOS a Research Infrastructure in solid Earth: open science and innovation, Ann. Geophys., 65, 2, 1, doi:10.4401/ag-8835.

Cocco, M., C. Freda, K. Atakan., D. Bailo et al. (2022). The EPOS Research Infrastructure: a federated approach to integrate solid Earth science data and services [Application/pdf], 15, doi:10.3929/ETHZ-B-000549762.

Cocco, M., R. Paciello, D. Bailo, D. M. Locati et al. (2025). The ethical dimension of sharing solid Earth Science data. J. Geoeth. Social Geosci., 2 (Special Issue), 1-30, doi:10.13127/JGSG-64.

Di Toro, G., A. R. Niemeijer, A. Tripoli, S. Nielsen et al. (2010). From field geology to earthquake simulation: A new state-of-The-art tool to investigate rock friction during the seismic cycle (SHIVA), Rendiconti Lincei, 21, S1, 95-114, doi:10.1007/s12210-010-0097-x.

Di Toro, G., A. Niemeijer and S. Nielsen (2025). Shiva data and metadata (s0050) [Data set]. Zenodo., Retrieved from https://zenodo.org/records/15783658.

DzIewonski, A. M. (1994). The FDSN: history and objectives, Ann. Geophys., 37, 5, 26, doi:10.4401/ag-4191.

Elger, K., G. ter Maat, R. Caldeira, C. Cimarelli et al. (2022). The EPOS Multi-Scale Laboratories: A FAIR Framework for stimulating Open Science practice across European Earth Sciences Laboratories, Ann. Geophys., 65, 3, doi:10.4401/ag-8790.

Fisher, A. T., T. Urabe., A. Klaus and the Expedition 301 Scientists (Eds.) (2005). Proceedings of the IODP, 301, 301, Integrated Ocean Drilling Program. doi:10.2204/iodp.proc.301.2005.

Haak, L. L., M. Fenner, L. Paglione, E. Pentz and H. Ratner (2012). ORCID: a system to uniquely identify researchers, Learned Publishing, 25, 4, 259-264. doi:10.1087/20120404.

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

Klump, J., K. Lehnert, D. Ulbricht, A. Devaraju, K. Elger et al. (2021). Towards Globally Unique Identification of Physical Samples: Governance and Technical Implementation of the IGSN Global Sample Number, Data Sci. J., 20, 1, doi:10.5334/dsj-2021-033.

Lehnert, K., Y. Su, C. H. Langmuir, B. Sarbas and U. Nohl (2000). A global geochemical database structure for rocks, Geochem., Geophys., Geosys., 1, 5, 1999GC000026, doi:10.1029/1999GC000026.

Mok, U. and M. Pec (2021, January 29). LAPS – A Digital Data Repository Workflow in Experimental Rock Deformation., doi:10.1002/essoar.10505991.1.

Noy, N. F. and D. L. McGuinness (2001). Ontology development 101: A guide to creating your first ontology., Retrieved from https://course.ccs.neu.edu/cs5100f11/resources/noy01.pdf.

Prent, A. M., D. C. Hezel, M. Klöcking, L. Wyborn et al. (2023). Innovating and networking global geochemical data resources through OneGeochemistry, Retrieved from https://eartharxiv.org/repository/view/5526/.

Rejiba, K., S. Korte-Kerzel and U. Kerzel (2024, September 23). openBIS tools for Materials Science and Engineering. Zenodo, doi:10.5281/zenodo.15731485.

Rejiba, K., S. Korte-Kerzel and U. Kerzel (2025, January 23). openBIS Schema Extension for research in Materials Science and Engineering. Zenodo, doi:10.5281/zenodo.15731469.

Rejiba, K., S.-H. Lee, C. Gasper, M. Freund, S. Korte-Kerzel and U. Kerzel (2025, November 3). Towards Defect Phase Diagrams: From Research Data Management to Automated Workflows. arXiv, doi:10.48550/arXiv.2511.01942.

Saleh Contell, K., K. Karlzén, M. Cocco, H. A. Pedersen et al. (2022). Long-term sustainability of a distributed RI: the EPOS case., Ann. Geophys., 65, 2, DM209, doi:10.4401/ag-8786.

Scholz, C. H. (1998). Earthquakes and friction laws, Nature, 391, 6662, 37-42, doi:10.1038/34097.

Siemer, N., L. Huber, Memebers SFB1394, S. Korte-Kerzel, S. and J. Neugebauer (2022). Metadata schemes for materials science data (Version 1) [Data set]. Zenodo, doi:10.5281/zenodo.6513745.

Stahlman, G. R. and I. Kouper (2026). Evolution of the "long-tail" concept for scientific data: An Annual Review of Information Science and Technology (ARIST) paper. J. Association Info. Sci. Technol., 77, 1, 3-22, doi:10.1002/asi.24967.

Tauxe, L., R. Shaar, L. Jonestrask, N. L. Swanson-Hysell et al. (2016). PmagPy: Software package for paleomagnetic data analysis and a bridge to the Magnetics Information Consortium (MagIC) Database., Geochem., Geophys., Geosys., 17, 6, 2450-2463, doi:10.1002/2016GC006307.

Tenopir, C., S. Allard, K. Douglass, A. U. Aydinoglu et al. (2011). Data Sharing by Scientists: Practices and Perceptions, PLoS ONE, 6,6,, e21101, doi:10.1371/journal.pone.0021101.

Townend, J. and M. D. Zoback (2000). How faulting keeps the crust strong. Geology, 28, 5, 399-402, doi:10.1130/0091-7613(2000)28%3C399:HFKTCS%3E2.0.CO;2.

Walker, J. D., B. Tikoff, J. Newman, R. Clark et al. (2019). StraboSpot data system for structural geology., Geosphere, 15, 2, 533-547, doi:10.1130/GES02039.1.

Wessels, R., G. ter Maat, E. D. Bello, L. Cacciola et al. (2022). Transnational Access to Research Facilities: an EPOS service to promote multi-domain Solid Earth Sciences in Europe, Ann. Geophys., 65, 2, DM214-DM214, doi:10.4401/ag-8768.

Wibberley, C. A. J., G. Yielding and G. Di Toro (2008). Recent advances in the understanding of fault zone internal structure: a review, Geol. Soc., London, Special Pub, 299,1, 5-33, doi:10.1144/SP299.2.

Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data, 3, 1, 160018, doi:10.1038/sdata.2016.18

***CORRESPONDING AUTHOR: Stefano ARETUSINI**,
Istituto Nazionale di Geofisica e Vulcanologia, Sezione Roma 1, Roma, Italy
e-mail: stefano.aretusini@ingv.it