Helmholtz Metadata Collaboration | Conference 2022

Report of Contributions

Contribution ID: **3** Type: **not specified**

Welcome notes

Wednesday 5 October 2022 09:00 (10 minutes)

Please assign your poster to one of the following keywords.

In addition please add keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Please specify ”other” (stakeholder)

Presenter: LORENZ, Sören (GEOMAR Helmholtz Centre for Ocean Research Kiel)

Session Classification: Welcome & Introduction

Contribution ID: **4** Type: **not specified**

Introduction to HMC

Wednesday 5 October 2022 09:10 (10 minutes)

Please assign your poster to one of the following keywords.

In addition please add keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Please specify ”other” (stakeholder)

Presenter: LORENZ, Sören (GEOMAR Helmholtz Centre for Ocean Research Kiel)

Session Classification: Welcome & Introduction

Contribution ID: **5** Type: **not specified**

HMC Impulse ”How FAIR is my data? Benefits and pitfalls of quantitative assessment of FAIRness.”

Wednesday 5 October 2022 10:20 (15 minutes)

Publishing data in a FAIR [1] way is already part of good scientific practice. While institutional policy as well as funding and publishing guidelines support this, scientist, technicians, and data stewards struggle to realize it when handling their research data. The reason is that the FAIR principles are high level principles and guidelines rather than concrete implementations. This is one of the key missions of HMC: support the Helmholtz community in making their data FAIR in an easy and comparable way. Developing a sustainable strategy for this requires a detailed understanding of practices, strengths, and deficiencies with respect to applying each of the FAIR principles. Here, tools that assess data FAIRness in comparison to a set of specific implementations in a quantitative fashion can help. When handling a dataset, such measures can aid the understanding of how FAIR a dataset actually is, as well as how to improve its FAIRness.
In this Blitzlicht-Talk, HMC Hub Matter and Hub Information will jointly present insights, benefits, and pitfalls from applying and further developing such metrics. For this we used the F-UJI tool [2,3], a python-based development by the FAIRsFAIR project, in two complementary projects.
In a “top-down”approach, we evaluate data repositories based on the data contained. The analyzed results are then used towards informing infrastructural development towards improving data FAIRness.
In a second, “bottom-up”approach, data publications from individual research centers or specific fields are evaluated with F-UJI. The results are gathered and visualized in an interactive pilot dashboard. This helps to identify and quantify the usage of repositories by Helmholtz‘s research communities as well as to better support the development of relevant infrastructure for FAIR data practices.
We discuss our experience from these automatic FAIR assessment approaches and compare them to complementary insights from a manual FAIR assessment of a particular data pipeline [4] using the FAIR Data Maturity Model [5]. We discuss future plans for metric development and the potential use of such metrics in user-sided tooling.

Please assign your poster to one of the following keywords.

In addition please add keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Please specify ”other” (stakeholder)

Presenters: KUBIN, Markus (HMC, HZB); VIDEGAIN BARRANCO, Pedro (Forschungszentrum Jülich)

Session Classification: Session

Contribution ID: **6** Type: **not specified**

Project eFAIRs: current status and next steps

Wednesday 5 October 2022 10:35 (15 minutes)

The seismological community promotes since decades standardisation of formats and services as well as open data policies which are making easy data exchange an asset for this community. Thus, data is made perfectly Findable and Accessible as well as Interoperable and Reusable with enhancements expected for the latter two. The strict and technical domain specific standardisation may complicate the sharing of more exotic data within the domain itself as well as hinder interoperability throughout the earth science community. Within eFAIRs, leveraging on the know-how of the major OBS park operators and seismological data curators within the Helmholtz association, we aim at facilitating integration of special datasets from the ocean floor enhancing interoperability and reusability.
To achieve this goal, in close collaboration with AWI and Geomar, supported by IPGP, the seismological data archive of the GFZ has created special workflows for OBS data curation. In particular, with close interaction with AWI, new datasets have been archived defining a new workflow which is being translated into guidelines for the community. Domain specific software have been modified to allow OBS data inclusion with specific additional metadata. Among these metadata also persistent identifiers of the instruments in use have been included for the first time from the AWI sensor information system. Next steps are going to enlarge the portfolio of keywords and standard vocabularies in use to facilitate data discovery from scientists of different domains. Finally we plan to adopt the developed workflows for OBS data management.

Please assign your poster to one of the following keywords.

In addition please add keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Please specify ”other” (stakeholder)

Primary author: STROLLO, Angelo (GFZ)

Co-author: HILLMANN, L.

Presenter: STROLLO, Angelo (GFZ)

Session Classification: Session

Contribution ID: **7** Type: **not specified**

Project HERMES: Automated FAIR4RS software publication with HERMES

Wednesday 5 October 2022 10:50 (15 minutes)

Software as an important method and output of research should follow the RDA "FAIR for Research Software Principles". In practice, this means that research software, whether open, inner or closed source, should be published with rich metadata to enable FAIR4RS.
For research software practitioners, this currently often means following an arduous and mostly manual process of software publication. HERMES, a project funded by the Helmholtz Metadata Collaboration, aims to alleviate this situation. We develop configurable, executable workflows for the publication of rich metadata for research software, alongside the software itself.
These workflows follow a push-based approach: they use existing continuous integration solutions, integrated in common code platforms such as GitHub or GitLab, to harvest, unify and collate software metadata from source code repositories and code platform APIs. They also manage curation of unified metadata, and deposits on publication platforms. These deposits are based on deposition requirements and curation steps defined by a targeted publication platform, the depositing institution, or a software management plan.
In addition, the HERMES project works to make the widely-used publication platforms InvenioRDM and Dataverse "research software-ready", i.e., able to ingest software publications with rich metadata, and represent software publications and metadata in a way that supports findability, assessability and accessibility of the published software versions.
Beyond the open source workflow software, HERMES will openly provide templates for different continuous integration solutions, extensive documentation, and training material. Thus, researchers are enabled to adapt automated software publication quickly and easily.
In this presentation, we provide an overview of the project aims, its current status, and an outlook on future development.

Please assign your poster to one of the following keywords.

In addition please add keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Please specify "other" (stakeholder)

Primary author: DRUSKAT, Stephan (German Aerospace Center (DLR))

Presenter: DRUSKAT, Stephan (German Aerospace Center (DLR))

Session Classification: Session

Contribution ID: **8** Type: **not specified**

Project HELIPORT: The Integrated Research Data Lifecycle of the HELIPORT Project

Wednesday 5 October 2022 11:05 (15 minutes)

The HELIPORT project aims to make the components or steps of the entire life cycle of a research project at the Helmholtz-Zentrum Dresden-Rossendorf (HZDR) and the Helmholtz-Institute Jena (HIJ) discoverable, accessible, interoperable and reusable according to the FAIR principles. In particular, this data management solution deals with the entire lifecycle of research experiments, starting with the generation of the first digital objects, the workflows carried out and the actual publication of research results. For this purpose, a concept was developed that identifies the different systems involved and their connections. By integrating computational workflows (CWL and others), HELIPORT can automate calculations that work with metadata from different internal systems (application management, Labbook, GitLab, and further). This presentation will cover the first year of the project, the current status and the path taken so far in the life cycle of the project.

Please assign your poster to one of the following keywords.

In addition please add keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Please specify ”other” (stakeholder)

Presenter: KNODEL, Oliver (Helmholtz-Zentrum Dresden-Rossendorf)

Session Classification: Session

Contribution ID: **9** Type: **not specified**

RO-Crate: packaging metadata love notes into FAIR Digital Objects

Wednesday 5 October 2022 09:25 (45 minutes)

The Helmholtz Metadata Collaboration aims to make the research data [and software] produced by Helmholtz Centres FAIR for their own and the wider science community by means of metadata enrichment [1]. Why metadata enrichment and why FAIR? Because the whole scientific enterprise depends on a cycle of finding, exchanging, understanding, validating, reproducing), integrating and reusing research entities across a dispersed community of researchers.

Metadata is not just "a love note to the future"[2], it is a love note to today's collaborators and peers. Moreover, a FAIR Commons must cater for the metadata of all the entities of research – data, software, workflows, protocols, instruments, geo-spatial locations, specimens, samples, people (well as traditional articles) –and their interconnectivity. That is a lot of metadata love notes to manage, bundle up and move around. Notes written in different languages at different times by different folks, produced and hosted by different platforms, yet referring to each other, and building an integrated picture of a multi-part and multi-party investigation. We need a crate!

RO-Crate [3] is an open, community-driven, and lightweight approach to packaging research entities along with their metadata in a machine-readable manner. Following key principles - "just enough"and "developer and legacy friendliness - RO-Crate simplifies the process of making research outputs FAIR while also enhancing research reproducibility and citability. As a self-describing and unbounded "metadata middleware"framework RO-Crate shows that a little bit of packaging goes a long way to realise the goals of FAIR Digital Objects (FDO)[4], and to not just overcome platform diversity but celebrate it while retaining investigation contextual integrity.

In this talk I will present the why, and how Research Object packaging eases Metadata Collaboration using examples in big data and mixed object exchange, mixed object archiving and publishing, mass citation, and reproducibility. Some examples come from the HMC, others from EOSC, USA and Australia, and from different disciplines.

Metadata is a love note to the future, RO-Crate is the delivery package.

Please assign your poster to one of the following keywords.

In addition please add keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Please specify "other" (stakeholder)

Primary author: Prof. GOBLE, Carole (The University of Manchester)

Presenter: Prof. GOBLE, Carole (The University of Manchester)

Session Classification: Keynote I

Contribution ID: **10** Type: **not specified**

HMC Impulse ”A basic Helmholtz Kernel Information Profile for machine-actionable FAIR Digital Objects”

Wednesday 5 October 2022 12:05 (15 minutes)

To reach the declared goal of the Helmholtz Metadata Collaboration Platform, making the depth and breadth of research data produced by Helmholtz Centres findable, accessible, interoperable, and reusable (FAIR) for the whole science community, the concept of FAIR Digital Objects (FAIR DOs) has been chosen as top-level commonality across all research fields and their existing and future infrastructures.
Over the last years, not only by the Helmholtz Metadata Collaboration Platform, but on an international level, the roads towards realizing FAIR DOs has been paved more and more by concretizing concepts and implementing base services required for realizing FAIR DOs, e.g., different instances of Data Type Registries for accessing, creating, and managing Data Types required by FAIR DOs and technical components to support the creation and management of FAIR DOs: The Typed PID Maker providing machine actionable interfaces for creating, validating, and managing PIDs with machine-actionable metadata stored in their PID record, or the FAIR DO testbed, currently evolving into the FAIR DO Lab, serving as reference implementation for setting up a FAIR DO ecosystem.
However, introducing FAIR DOs is not only about providing technical services, but also requires the definition and agreement on interfaces, policies, and processes.
A first step in this direction was made in the context of HMC by agreeing on a Helmholtz Kernel Information Profile. In the concept of FAIR DOs, PID Kernel Information is key to machine actionability of digital content. Strongly relying on Data Types and stored in the PID record directly at the PID resolution service, PID Kernel Information is allowed to be used by machines for fast decision making.
In this session, we will shortly present the Helmholtz Kernel Information Profile and a first demonstrator allowing the semi-automatic creation of FAIR DOs for arbitrary DOIs accessible via the well-known Zenodo repository.

Please assign your poster to one of the following keywords.

In addition please add keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Please specify ”other” (stakeholder)

Primary author: JEJKAL, Thomas

Presenter: JEJKAL, Thomas

Session Classification: Session

Contribution ID: **11** Type: **not specified**

Project FDO-5DI: Connecting seafloor and planetary surfaces: Approaching via an interoperable metadata description for imaging research data

Wednesday 5 October 2022 12:20 (15 minutes)

Imaging the environment is an essential and crucial component in spatial science. This concerns nearly everything between the exploration of the ocean floor and investigating planetary surfaces. In and between both domains, this is applied at various scales –from microscopy through ambient imaging to remote sensing –and provides rich information for science. Due to recent the increasing number data acquisition technologies, advances in imaging capabilities, and number of platforms that provide imagery and related research data, data volume in nature science, and thus also for ocean and planetary research, is further increasing at an exponential rate. Although many datasets have already been collected and analyzed, the systematic, comparable, and transferable description of research data through metadata is still a big challenge in and for both fields. However, these descriptive elements are crucial, to enable efficient (re)use of valuable research data, prepare the scientific domains e.g. for data analytical tasks such as machine learning, big data analytics, but also to improve interdisciplinary science by other research groups not involved directly with the data collection. In order to achieve more effectiveness and efficiency in managing, interpreting, reusing and publishing imaging data, we here present a project to develop interoperable metadata recommendations in the form of FAIR [1] digital objects (FDOs) [2] for 5D (i.e. x, y, z, time, spatial reference) imagery of Earth and other planet(s). An FDO is a human and machine-readable file format for an entire image set, although it does not contain the actual image data, only references to it through persistent identifiers (FAIR marine images [3]). In addition to these core metadata, further descriptive elements are required to describe and quantify the semantic content of imaging research data. Such semantic components are similarly domain-specific but again synergies are expected between Earth and planetary research. We here present the current status of the project, with the specific tasks on joint metadata description of planetary and oceanic data.

Please assign your poster to one of the following keywords.

In addition please add keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Please specify ”other” (stakeholder)

Primary author: NASS, Andrea (DLR)

Presenter: NASS, Andrea (DLR)

Session Classification: Session

Contribution ID: **12** Type: **not specified**

Project FAIR WISH: FAIR Workflows to establish IGSN for Samples in the Helmholtz Association

Wednesday 5 October 2022 12:35 (15 minutes)

Physical samples or specimen are often at the beginning of the "research chain"as they are the source for many data described in scholarly literature. The International Generic Sample Number (IGSN) is a globally unique and persistent identifier (PID) for physical samples and collections with discovery function in the internet. IGSNs enable to directly link data and publications with samples they originate from and thus close the last gap in the full provenance of research results. The modular IGSN metadata schema has a small number of mandatory and recommended metadata elements that can be individually extended with discipline-specific elements.

Based on three use cases that represent all states of digitisation - from individual scientists, collecting sample descriptions in their field books to digital sample management systems fed by an app that is used in the field - FAIR WISH will (1) develop standardised and discipline specific IGSN metadata schemes for different sample types from the Earth and Environment Sciences, (2) develop workflows to generate machine-readable IGSN metadata from different states of digitisation, (3) develop workflows to automatically register IGSNs and (4) prepare the resulting workflows for further use in the Earth Science community.

After investigating and identifying controlled linked-data vocabularies that can be included in our metadata schema, we recently have published the first data description template that includes new fields for biological and water samples. The template can be used by researchers to provide their sample descriptions and will serve as basis for semi-automated metadata generation.

Primary author: ELGER, Kirsten (GFZ)

Presenter: ELGER, Kirsten (GFZ)

Session Classification: Session

Contribution ID: **13** Type: **not specified**

HMC Impulse ”Use cases in HMC - from generation to reuse of data”

Thursday 6 October 2022 09:10 (15 minutes)

We present three use cases which showcase methods of providing a detailed metadata description with the goal of increasing the reusability of data.
irst, Hub Energy presents a photovoltaic system which required ontology development and the implementation of data models based on standards like IEC 61850 [1] or SensorML [2] as well as on FAIR Digital Objects (FDO) [3]. The backend was realized using the Metastore [4] software from the Fair Data Commons while a FDO browser was implemented for visualization which offers a cascading search for metadata and application data.
In a second use case of Hub Energy, time series data of the energy consumption of the buildings on KIT’s Campus North are described by automatically generated RO-Crates [5]. This allows energy researchers to use these time series data without any knowledge about the structure of the database and provides a case study on working with RO-Crate technology.

The third use case is provided by Hub Matter, in the research field of high energy physics, and shows the optimization of a typical data set for data publication. To increase FAIRness of the distributed file set, (meta)data is (i) enriched by metadata, (ii) converted to a machine- as well as human-readable format and (iii) linked to a central file to create scientific context. By abstracting from community-specific details these measures can serve as a general approach to make data publishable.
The variety of use cases presented provides a menu of technologies and approaches implemented in diverse contexts to enhance the reusability of data, along with general advice for anyone looking to do the same.

Please assign your poster to one of the following keywords.

In addition please add keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Please specify ”other” (stakeholder)

Presenters: GUENTHER, Gerrit (Helmholtz-Zentrum Berlin); SCHWEIKERT, Jan (KIT)

Session Classification: Session

Contribution ID: **14** Type: **not specified**

Project AutoPeroSol: Towards automatic data management and a common ontology for perovskite solar cell device data

Thursday 6 October 2022 09:25 (15 minutes)

A general photovoltaic device and materials data base compliant with FAIR principles is expected to greatly benefit research and development of solar cells. Because data are currently heterogeneous in different labs working on a variety of different materials and cell concepts, database development should be accompanied by ontology development. Based on a recently published literature database for perovskite solar cells, we have started an ontology for these devices and materials which could be extended to further photovoltaic applications. In order to facilitate data management at the lab scale and to allow easy upload of data and metadata to the database, electronic lab notebooks customized for perovskite solar research are developed in cooperation with the NFDI-FAIRmat project. Current status and challenges will be discussed.

Please assign your poster to one of the following keywords.

In addition please add keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Please specify ”other” (stakeholder)

Primary author: UNOLD, Thomas (HZB)

Presenter: UNOLD, Thomas (HZB)

Session Classification: Session

Contribution ID: **15** Type: **not specified**

Project MetaMoSim: Generic metadata management for reproducible high-performance-computing simulation workflows

Thursday 6 October 2022 09:40 (15 minutes)

Modern science is to a vast extent based on simulation research. With the advances in high-performance computing (HPC) technology, the underlying mathematical models and numerical workflows are steadily growing in complexity.
This complexity gain offers a huge potential for science and society, but simultaneously constitutes a threat for the reproducibility of scientific results. A main challenge in this field is the acquisition and organization of the metadata describing the details of the numerical workflows, which are necessary to replicate numerical experiments, and to explore and compare simulation results. In the recent past, various concepts and tools for metadata handling have been developed in specific scientific domains. It remains unclear to what extent these concepts are transferable to HPC based simulation research, and how to ensure interoperability in the face of the diversity of simulation based scientific applications. This project aims at developing a generic, cross-domain metadata management framework to foster reproducibility of HPC based simulation science, and to provide workflows and tools for an efficient organization, exploration and visualization of simulation data. Within the project, we so far did a review of existing approaches from different fields. A plethora of tools around metadata handling and workflows have been developed in the past years. We identified tools and formats like the odML that are useful for our work. The metadata management framework will address all components of simulation research and the corresponding metadata types, including model description, model implementation, data exploration, data analysis, and visualization. We have now developed a general concept to track, store and organize metadata. Next, the required tools within the concept will be developed such that they are applicable both in the Computational Neuroscience and Earth and Environmental Science.

Please assign your poster to one of the following keywords.

In addition please add keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Please specify ”other” (stakeholder)

Primary author: THOBER, Stephan (Dept. Computational Hydrosystems, Helmholtz-Centre for Environmental Research, Leipzig, Germany)

Presenter: THOBER, Stephan (Dept. Computational Hydrosystems, Helmholtz-Centre for Environmental Research, Leipzig, Germany)

Session Classification: Session

Contribution ID: **16** Type: **not specified**

HMC Impulse ”HMC initiatives towards interoperable semantics in research”

Thursday 6 October 2022 10:05 (15 minutes)

Scientific technology, the supporting infrastructure and the resulting data are highly complex and extremely diverse. The work of the past decades has achieved digitization of many aspects in research, such as exeriments, instrumentation or the publishing process. However these individual parts mostly remain “digitized islands”and, so far, we ar lacking a systematic, broad and interoperable connection between them. Here, both formalization and standardisation of data descriptions within and across research fields, i.e. research data interoperability, remain a major challenge.
A core HMC action is to support interoperabiltiy within the Helmholtz digital ecosystem, and to ensure its alignment with global technology and standards at the same time. Both of these tasks require cooperation on many levels, ranging from the level of domain scientists and their research data, the level of data stewardship and knowledge engineering to the infrastructural and institutional level.
In this talk we will present HMC initiatives, developments and services that are all working towards an interoperable Helmholtz digital ecosystem: At the application and domain level, we are working with the electron microscopy community towards homogenized and interoperable semantic descriptions in these fields. At the level of data stewardship and knowledge engineering, HMC provides services such as our lightweight PID service PIDA and develops the “Helmholtz digitization ontology”(HDO). Once released, HDO will provide a harmonized, formal and machine-actionable understanding of the key concepts around digital dataspaces. At the infrastructural and institutional level, we are developing a Helmholtz Knowledge Graph. We will present first steps and a sketch about how this Knoledge Graph will link Helmholtz infrastructures to create a system that allows research data to be found and exchanged, both within the Helmholtz association and with the globally operating systems.

Please assign your poster to one of the following keywords.

In addition please add keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Please specify ”other” (stakeholder)

Primary author: HOFMANN, Volker

Presenter: HOFMANN, Volker

Session Classification: Session

Contribution ID: **17** Type: **not specified**

Project ADVANCE: Advanced metadata standards for biodiversity survey and monitoring data: Supporting of research and conservation

Thursday 6 October 2022 10:20 (15 minutes)

In an ever-changing world, field surveys, inventories and monitoring data are essential for prediction of biodiversity responses to global drivers such as land use and climate change. This knowledge provides the basis for appropriate management. However, field biodiversity data collected across terrestrial, freshwater and marine realms are highly complex and heterogeneous. The successful integration and re-use of such data depends on how FAIR (Findable, Accessible, Interoperable, Reusable) they are. ADVANCE aims at underpinning rich metadata generation with interoperable metadata standards using semantic artefacts. These are tools allowing humans and machines to locate, access and understand (meta) data, and thus facilitating integration and reuse of biodiversity monitoring data across terrestrial, freshwater and marine realms. To this end, we revised, adapted and expanded existing metadata standards, thesauri and vocabularies. We focused on the most comprehensive database of biodiversity monitoring schemes in Europe (DaEuMon) as the base for building a metadata schema that implements quality control and complies with the FAIR principles. In a further step, we will use biodiversity data to test, refine and illustrate the strength of the concept in cases of real use. ADVANCE thus complements semantic artefacts of the Hub Earth & Environment and other initiatives for FAIR biodiversity research, enabling assessments of the relationships between biodiversity across realms and associated environmental conditions. Moreover, it will facilitate future collaborations, joint projects and data-driven studies among biodiversity scientists of the Helmholtz Association and beyond.

Please assign your poster to one of the following keywords.

In addition please add keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Please specify ”other” (stakeholder)

Presenters: GRIMM-SEYFARTH, Annegret (UFZ); SILVA MENGER, Juliana (UFZ, AWI)

Session Classification: Session

Contribution ID: **18** Type: **not specified**

Project MetaMap3: Metadata generation, enrichment and linkage across the three domains health, environment and earth observation: the MetaMap³ project

Thursday 6 October 2022 10:35 (15 minutes)

Digital metadata solutions for epidemiological cohorts are lacking since most schemas and standards in the Health domain are clinically oriented and cannot be directly transferred. In addition, the environment plays an increasingly important role for human health and efficient linkage with the multitude of environmental and earth observation data is crucial to quantify human exposures. There are however currently no harmonized metadata standards for the different areas, so they cannot be merged routinely. Therefore, we aim to compile machine-readable and interoperable metadata schemas for exemplary data of our three domains Health (HMGU), Earth & Environment (UFZ), and Aeronautics, Space & Transport (DLR).
We will present our data use cases (HMGU: GINI/LISA cohort; UFZ: drought monitor; DLR: land cover), their current metadata formats and our strategy for metadata compilation, enrichment and mapping. UFZ and DLR will converge their metadata to the standard ISO 19115: Geographic Metadata Information. For HMGU, we reviewed several metadata standards for health data (e.g. CDISC ODM, Snomed CT, HL7 FIHR) and started to upload our metadata to the NFDI4health StudyHub, an inventory of German health studies on COVID-19 which is based on the Maelstrom catalogue. In addition, we have developed a workflow to transform base cohort information in an ISO 19115 compliant manner. The respective metadata sheet increases accessibility to researchers from other domains without exposing sensitive information about participants'data.
The metadata mapping will be performed by location (spatial coverage) and date (time coverage) within GeoNetwork, a catalog application that we are currently setting up in a testing environment. We aim to have a server version ready by the end of the project that can be augmented with additional metadata from our domains, but also from other fields, to facilitate interdisciplinary research.

Please assign your poster to one of the following keywords.

In addition please add keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Please specify "other" (stakeholder)

Primary author: WOLF, Kathrin (Helmholtz Zentrum München)

Presenter: WOLF, Kathrin (Helmholtz Zentrum München)

Session Classification: Session

Contribution ID: **20** Contribution code: **2-31** Type: **Poster**

Technology neutral provenance storage & sharing

Provenance is one of the requirements for reusable data (see FAIR principles). There are data formats, which store data and provenance (metadata) easily together like hdf5, data package, research objects and others. Nevertheless, these are not applicable to all data and all use cases. Therefore, provenance/metadata management systems are often used. Unfortunately, there are at least two problems with such systems: 1. Maintenance effort (in various forms like costs, organizational overhead, vendor & technology lock-in and therefore slowed down development) with respect to long term data reuse (decades) and 2. Incompatible IT landscapes between different data sharing stakeholders which will not be synchronized (due to costs, different IT policies, time, …) and therefore block data/provenance exchange.

We developed a simple concept for storing & sharing provenance between different stakeholders along with data. The so-called provenance container emphasizes a "provenance first" approach and consists of the unchanged data and an additional provenance description. Provenance is provided using common standards in extendable W3C prov model, serialization as json plain text and identified in a content addressable way. We use hash sums to reference from provenance to data without the need for additional reference systems or data formats. The provenance trace created with the container is effectively unforgeable once shared with other stakeholders. Provenance container need no storage and sharing requirements different than the minimal requirements for the data itself. Provenance container are technology neutral due to the simple design with only standard tools like hash sums, json and plain text. Human readability is given due to plain text and json as information encoding.

This poster will highlight and show case the main attributes of provenance containers, its pros and cons and how to use it for easy data storage & sharing between different stakeholders.

Please assign your poster to one of the following keywords.

Processes/Policies

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Producer

Please specify "other" (stakeholder)

In addition please add keywords.

provenance
data sharing
data reuse

Primary author: DRESSEL, Frank

Presenter: DRESSEL, Frank

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **21** Contribution code: **2-06** Type: **Poster**

Metadata for a FAIRer World

Metadata plays a key role in the scientific publication process. It is only through metadata and identifiers that each contribution, from research data to article publication and beyond, becomes findable, accessible, interoperable and reusable. The digitization of scholarly communication allows the creation of metadata locally or in a distributed manner, and global exchange, enabled by machine readability. Persistent identifiers (PIDs) and their metadata are the backbone of scholarly communications, since only through them can the promise of sustainable access to research information be realized. However, despite years of steadily improving metadata management capabilities, the completeness of metadata falls far short of its potential.

This poster will highlight the importance of metadata in scholarly communications and the need for collaborative, community efforts to improve its creation and curation. The relevance of metadata for PIDs in general and Digital Object Identifiers (DOIs) in particular will be addressed. The structure and status quo of DataCite DOI metadata will be presented through analyses of the DataCite metadata schema and the use of metadata fields.

Please assign your poster to one of the following keywords.

Standards

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify ”other” (stakeholder)

In addition please add keywords.

Metadata
Persistent Identifiers
Research Infrastructure

Primary authors: MEJIAS, Gabriela (DataCite); VIERKANT, Paul (DataCite)

Presenters: MEJIAS, Gabriela (DataCite); VIERKANT, Paul (DataCite)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **22** Contribution code: **2-05** Type: **Poster**

Annotating Humanities Research Data

Annotation is one of the oldest cultural techniques of mankind. While in past centuries pen and paper were the means of choice to add annotations to a source, this activity has increasingly shifted to the digital world in recent years. With the W3C recommendation 'Web Annotation Data Model', a powerful tool has been available since 2017 to model annotations in a wide variety of disciplines and to enable cross-disciplinary analysis.

In this poster, we would like to give an insight into our annotation infrastructure, which is in use in three humanities research projects. The focus is on a custom-developed annotation server with RDF backend (fully compliant to the 'Web Annotation Protocol') as well as our annotation interfaces. The interaction of these components with each other but also with other infrastructure components such as a research data repository or a vocabulary editor as well as the daily work of researchers with these components will be illustrated.

Special attention will be paid to the modeling of annotations in our different use cases. Examples range from labeling of logical diagrams in medieval Aristotle manuscripts, to the analysis of metaphors in religious meaning-making, to the capture of particular Hebrew letter variations in Torah scrolls. The discussion of similarities and differences in these use cases holds great potential for transferability and fruitfulness in further scientific disciplines and is thus a main part of our contribution.

Please assign your poster to one of the following keywords.

Standards

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify "other" (stakeholder)

In addition please add keywords.

WADM
SKOS
Annotation infrastructure

Primary authors: TONNE, Danah (Karlsruhe Institute of Technology); ERNST, Felix; GÖTZELMANN, Germaine; FRANK, Laura; TÖGEL, Philipp; Dr JHA, Vandana

Presenters: TONNE, Danah (Karlsruhe Institute of Technology); GÖTZELMANN, Germaine; FRANK, Laura; TÖGEL, Philipp; Dr JHA, Vandana

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **23** Contribution code: **1-14** Type: **Poster**

A metadata and data entry and editing tool using ontologies for knowledge graph creation

Making research reproducible and FAIR (Findable, Accessible, Interoperable, and Reusable) often requires more information than what is commonly published within scientific articles. There is a growing number of repositories for publishing additional material like data or code. However, articles are still at the center of most scientific work and thus efforts on gathering information which is important for reproducibility but not for the article itself are often only started at a later stage. This usually makes the collection more tedious, error-prone, and less comprehensive.

In order to lower the barrier for recording all relevant information directly when it is generated, we propose a design for a data and metadata entry and editing tool. It should allow researchers to create metadata for the files and assets which they already have and offer a possibility for structured entry of new data. To support consistent (meta)data entry over time, the user will be able to create forms which can enforce comprehensiveness and correctness. Furthermore, data FAIRness is supported through the automated usage of established ontologies for (meta)data annotation. This will be done by a background process so the user isn´t involved in these technologies. Nevertheless the tool will grant further possibilities to those who are aware of ontologies used in their domain. Resources can be referenced consistently across (meta)data sets of many stakeholders through identifiers provided by central sources. In conjunction, the usage of these common identifiers and ontologies forms large knowledge graphs of the data recorded with our tool.

This contribution will be a discussion of the various components of such a tool, potentially used metadata standards, possible variations, important features, and most relevant: How it can be useful for your work!

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify ”other” (stakeholder)

Metadata Developer for HMC

In addition please add keywords.

Data Entry Harmonization, Knowledge Graphs

Primary authors: STEINMEIER, Leon (Helmholtz Institute Freiberg); Prof. VAN DEN BOOGAART, Karl Gerald (HZDR/HIF); Dr RAU, Florian (HZDR/HIF); SCHALLER, Theresa (HZDR/HIF)

Presenter: STEINMEIER, Leon (Helmholtz Institute Freiberg)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **24** Contribution code: **2-33** Type: **Poster**

Tracking large-scale simulations through unified metadata handling

Simulation is an essential pillar of knowledge generation in science. The numerical models used to describe, predict, and understand real-world systems are typically complex. Consequently, applying these models by means of simulation often poses high demands on computational resources, and requires high-performance computing (HPC) or other dedicated hardware architectures. Metadata describing the details of a numerical experiment arise at all stages of the simulation process: the conceptual description of the model, the model implementation, and the tools and machines used to run the simulation. Capturing these metadata and provenance information along the processing chain is a vital requirement for several purposes, e.g. reproducibility, benchmarking and validation, assessment of the reliability of the simulations, and data exploration[1,2]. The ability to search, share, and evaluate metadata and provenance traces from heterogeneous simulations and environments is a major challenge in provenance-driven analysis. The availability of a common metadata framework, which can be adopted by scientists from different scientific domains, would foster the meta-analysis of HPC simulation workflows[3]. Here, we develop a metadata management framework for generic HPC-based simulation research comprising concepts and tools for efficiently generating, organizing, and exploring metadata along a given simulation workflow. The derived solutions cope with the modularity and flexibility demands of rapidly progressing science and are applicable to diverse research fields. As a proof of concept, we will apply these solutions to use cases from environmental research and computational neuroscience.

References:
1. Guilyardi, E., et. al. (2013) doi: 10.1175/BAMS-D-11-00035.1
2. Manninen, T., et. al. (2018) doi: 10.3389/fninf.2018.00020
3. Ivie, P., & Thain, D. (2018) doi: 10.1145/3186266

Acknowledgements:
The authors would like to thank Jan Bumberger, Helen Kollai, Michael Denker, Rainer Stotzka, Guido Trensch, and Stefan Sandfeld for ongoing fruitful discussion. This project was funded by Helmholtz Metadata Collaboration (HMC) ZT-I-PF-3-026, EU Grant 945539 (HBP), Helmholtz IVF Grant SO-092 (ACA), and Joint lab SMHB; compute time was granted by VSR computation grant JINB33, Jülich. The work was carried out in part within the HMC Hub Information at the Forschungszentrum Jülich.

Please assign your poster to one of the following keywords.

Processes/Policies

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Producer

Please specify ”other” (stakeholder)

In addition please add keywords.

Metadata-Framework, High-Performance-Computing, Simulation-Workflow, Reproducibility, Reusability

Primary authors: VILLAMAR, Jose (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany; RWTH Aachen University, Aachen, Germany); KELBLING, Matthias (Dept. Computational Hydrosystems, Helmholtz-Centre for Environmental Research, Leipzig, Germany); TERHORST, Dennis (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany); MORE, Heather (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany; Institute for Advanced Simulation (IAS-9), Jülich Research Centre, Jülich, Germany); TETZLAFF, Tom (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany); SENK, Johanna (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany); THOBER, Stephan (Dept. Computational Hydrosystems, Helmholtz-Centre for Environmental Research, Leipzig, Germany)

Presenter: VILLAMAR, Jose (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany; RWTH Aachen University, Aachen, Germany)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **25** Contribution code: **2-03** Type: **Poster**

ORCID in Germany - a project-driven success story

The Open Researcher and Contributor ID ORCID strives to enable transparent and trustworthy connections between researchers, their contributions, and their affiliations by providing a unique, persistent identifier for individuals to use as they engage in research, scholarship, and innovation activities. ORCID is therefore an essential piece of the puzzle for increasing the discoverability of researchers by disambiguating them from all the other researchers with the same or even a similar name and definitively connecting them with their research contribution metadata (e.g. their scholarly record). Furthermore, ORCID specifically addresses each of the FAIR findability principle components. The international non-profit organization ORCID, on which the initiative is based, already connects over 14 million persons worldwide with their research outcomes.
In order to promote the widespread implementation of ORCID at universities and non-university research institutions in Germany, the project "ORCID DE" was launched and is funded by the German Research Foundation (DFG).
With the development of claiming services for the national bibliography of the German National Library and the BASE (Bielefeld Academic Search Engine), one of the world's most voluminous search engines especially for academic web resources, ORCID DE provides central tools for shaping the distribution and quality of metadata of scholarly communication from Germany.
Furthermore the project fosters the interconnection of ORCID with other persistent identification (PID) systems and therefore contributes to more discoverability, accessibility, and visibility of research outcomes.
As coordinator of the project, the Helmholtz Open Science Office makes an important contribution to improving scientific information management in the context of Open Science

Further project partners of ORCID DE are Deutsche Nationalbibliothek, Universitätsbibliothek Bielefeld, DataCite - International Data Citation Initiative e. V. and Technische Informationsbibliothek (TIB) Hannover.

Please assign your poster to one of the following keywords.

Standards

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify "other" (stakeholder)

Open Science Officer

In addition please add keywords.

open, disambiguation, identifier, standard, access

Primary authors: SCHRADER, Antonia (Helmholtz Open Science Office); Dr PAMPEL, Heinz (Helmholtz Open Science Office)

Presenter: SCHRADER, Antonia (Helmholtz Open Science Office)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **26** Contribution code: **1-43** Type: **Poster**

Enabling open science practices in Helmholtz!

The Helmholtz Open Science Office embraces this mission (Enabling open science practices in Helmholtz!) since it was founded by the Helmholtz Association in 2005. It supports the Helmholtz Association as a service provider in shaping the cultural change towards open science. Furthermore, it promotes dialogue on open science within and beyond Helmholtz and regularly offers events on open science developments to provide impulses and guidelines for the Helmholtz Association.

In the area of open research data, the Helmholtz Open Science Office supports the Helmholtz Centers in developing policies and implementing corresponding practices for handling digital research data. Furthermore, the OS Office supports the discussion on the re-use and thus also reproducibility of research data in Helmholtz and the interdisciplinary exchange in this field.
The National Research Data Infrastructure (NFDI) and the European Open Science Cloud (EOSC) represent central infrastructures in this context. Numerous NFDI consortia are being implemented with substantial Helmholtz participation, and Helmholtz Centers are also actively involved in the implementation of the EOSC. The OS Office actively animates and supports these activities.
Helmholtz's participation in the international Research Data Alliance (RDA) and in Research Data Alliance Germany (RDA-DE) is also accompanied and coordinated by the Helmholtz Open Science Office. Moreover, the OS Office is active in third-party funded projects on open research data and persistent identifiers, such as re3data COREF and ORCID DE.
The work of the Helmholtz Open Science Office thus complements the developments of platforms in the Helmholtz Incubator and contributes to the utilization of the FAIR principles in Helmholtz, such as the Helmholtz Metadata Collaboration (HMC).

Please assign your poster to one of the following keywords.

Processes/Policies

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify "other" (stakeholder)

Open Science Officer

In addition please add keywords.

open, science, data, policy, FAIR

Primary authors: SCHRADER, Antonia (Helmholtz Open Science Office); PAMPEL, Heinz (Helmholtz Open Science Office); FERGUSON, Lea Maria (Helmholtz Association, Helmholtz Open Science Office); WEISWEILER, Nina (Helmholtz Open Science Office); BERTELMANN, Roland (Helmholtz Open Science Office)

Presenter: SCHRADER, Antonia (Helmholtz Open Science Office)

Track Classification: Postersession

Contribution ID: **27** Contribution code: **1-21** Type: **Poster**

Celebrating 10 Years of re3data –The Registry of Research Data Repositories

The year 2022 marks the 10th anniversary of the Registry of Research Data Repositories - re3data. The global index currently lists over 2,800 digital repositories across all scientific disciplines –critical infrastructures to enable the global exchange of research data. The openly accessible service is used by researchers and services worldwide. It provides extensive descriptions of repositories based on a detailed and publicly available metadata schema (https://www.doi.org/10.48440/RE3.010). Ingests in re3data are managed by an international Editorial Board. The editorial process includes a multi-stage review, adapting best practices in science. A team of research data professionals thoroughly analyzes the repositories and takes care of metadata completeness and quality.

The poster presents the growth, development, and accomplishments of re3data over the past 10 years that have resulted in re3data becoming the most comprehensive and largest information and metadata resource on research data repositories. The poster highlights and summarizes important activities like the initial projects funded by the German Research Foundation (DFG), the collaboration and merger with DataBib, and the partnership with DataCite. It also details collaborations like the EU Open Science Monitor, SNF, and FAIR-enabling measures together with AGU and the European Community. The poster addresses activities with different research communities in order to initiate intensive talks about best practices and experience made. Furthermore, it gives succint insights into the current projects re3data COREF (Community Driven Open Reference for Research Data Repositories / DFG funded) and FAIR Impact (EU funded).

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify ”other” (stakeholder)

In addition please add keywords.

registry, research data, repositories, standard

Primary authors: WEISWEILER, Nina Leonie (Helmholtz Open Science Office); BERTELMANN, Roland (Helmholtz Open Science Office)

Presenter: BERTELMANN, Roland (Helmholtz Open Science Office)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **28** Contribution code: **2-25** Type: **Poster**

Creation of a semantic interoperable application profile to deliver metadata on research datasets to Flanders Research Information Space.

Flanders Research Information Space a.k.a. the FRIS-portal is a research discovery platform hosted by the Flemish department Economy, Science and Innovation, where you can find information on publicly financed research in Flanders. The FRIS-portal operates as a regional metadata hub where you can search for metadata on researchers, research groups, projects, and publications and recently also datasets, patents and infrastructure. There are currently more than 60 research performing and funding institutions (universities, schools of higher education, strategic research centres and scientific institutions, Flemish funding organisations) that provide data to the platform in real-time using CERIF XML. The FRIS-portal is also used as a BI-tool to monitor and report on the Flemish policy on research and innovation.
In 2019, the Flemish government developed an Open Science policy to make publicly financed research data openly accessible in line with the principles of the European Open Science Cloud (EOSC): 'as open as possible as restricted as necessary'. A Flemish Open Science Board was set up with representatives from the knowledge institutions and tasked with creating a roadmap to implement this policy using a stakeholder approach. The FOSB established 5 KPI's around ORCID, Data Management Planning, Open Access, FAIR data and Open data to track progress in open science. The KPI's on FAIR data and Open data outline that datasets resulting from publicly funded research projects should become Findable, Accessible, Interoperable and Reusable, and openly available by default (except for legitimate opt-out reasons). To monitor these KPI's there was a need for a semantically harmonized application profile to deliver metadata on datasets to the FRIS-portal. This poster outlines the development of a Flemish application profile for research datasets that was developed by the Flemish Open Science Board (FOSB) Working Group Metadata & Standardization. The main challenge was to achieve semantic interoperability among a wide and diverse range of stakeholders.

In addition please add keywords.

Metadata; Semantic interoperability; Open Science

Please assign your poster to one of the following keywords.

Semantics

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify "other" (stakeholder)

Project member information modelling

Primary author: NEYENS, Evy (ECOOM-Hasselt)

Co-author: Dr VANCAUWENBERGH, Sadia (Hasselt University)

Presenter: NEYENS, Evy (ECOOM-Hasselt)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **29** Contribution code: **1-17** Type: **Poster**

MetaCook: FAIR Vocabularies Cookbook

One of the prerequisites for FAIR data publication is the use of FAIR vocabularies. Currently, tools for the collaborative composition of such vocabularies are missing. For this reason, a universal manual and software for user-friendly vocabulary assembly is being composed in the HMC-funded MetaCook project. The project includes 4 separate test cases from 4 labs across KIT and Hereon, which will help strengthen the software's universality and applicability to various domains.

The components described in MetaCook will be implemented in the form of multiple software tools. The first one, a Python-based web application called VocPopuli, is the entry point for domain experts. The software, whose first version is being developed at the time of writing, enables the collaborative definition, and editing of metadata terms. Additionally, it annotates each term, as well as the entire vocabulary, with the help of the PROV Data Model (PROV-DM) - a schema used to describe the provenance of a given object. Finally, it assigns a unique ID to each term in the vocabulary, as well as a hash-based ID the vocabulary itself.

The second software tool will facilitate the transformation of the vocabularies developed with the help of VocPopuli into ontologies. It will handle two distinct use cases –the from-scratch conversion of vocabularies into ontologies, and the augmentation of existing ontologies with the terms from a given thesaurus. Both software tools will be used by two semi-overlapping user groups: domain experts will input, edit, and discuss vocabulary terms in their area of interest, while vocabulary and ontology administrators will oversee the vocabulary creation, and ontology transformation.

Both the controlled vocabularies and the corresponding ontologies offer the possibility to enrich data documented in Electronic Laboratory Notebooks (ELNs). As the simplest solution, terms used within the ELN are linked to the IDs of the related vocabulary and ontology for an unambiguous definition. Additionally, an export of the defined schemes can be used to automatically create a structured form in the ELNs for documenting the described processes. The output from the developed tools will be exemplarily integrated into the ELNs Herbie and Kadi4Mat.

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Producer

Please specify ”other” (stakeholder)

In addition please add keywords.

Vocabularies, Ontology, ELN, FAIR, MaterialsScience

Primary authors: Dr GARABEDIAN, Nick (KIT); Prof. KLUSEMANN, Benjamin (Hereon); Mr BOCK, Frederic (Hereon); Prof. GREINER, Christian (KIT); Ms ESCHKE, Catriona (Hereon); Dr WIELAND, Florian (Hereon); Dr HELD, Martin (Hereon); Dr WEBER, Karlheinz (KIT); Mr BAGOV, Ilia (KIT)

Presenter: Dr GARABEDIAN, Nick (KIT)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **30** Contribution code: **1-11** Type: **Poster**

Use Cases and Tools in HMC Hub Energy

Five Helmholtz Centers are participating in the Research Field Energy, three of them are directly con-tributing to Hub Energy. To be well prepared for their supporting tasks in establishing a FAIR data ecosystem within the energy research community at Helmholtz, the team members of Hub Energy study relevant use cases and develop software tools in close cooperation with FAIR Data Commons. This poster presents four examples for this work: A photovoltaic system requires ontology develop-ment and data models based on standards like IEC 61850 or SensorML as well as on FAIR Digital Objects (FDO). In another use case, RO-Crates are automatically generated for data of the KIT Cam-pus North energy and water consumption. The aim is to study methods for a detailed metadata desc-iption in data publication processes. In the field of software development, an FDO browser offers cascading search for metadata and application data entities and a metadata editor supports users in creating and editing schemas and instances as well. The presented activities foster close contact between Hub Energy and Helmholtz energy researchers and, thus, essentially support the formation of a FAIR energy data management. Use cases feed technical details into the Hub's energy knowledge pool and they are also a nearly perfect training programme for the Hub personnel. In doing the presented software development work, deep insights into energy data landscapes and an im-proved sense for user requirements are induced, even if in the end more elaborated and harmonized solutions from FAIR Data Commons may be adopted.

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify "other" (stakeholder)

HMC Staff Hub Energy

In addition please add keywords.

Use Cases

Primary authors: Dr BALLANI, Felix (HZDR); SCHWEIKERT, Jan (KIT); Dr STUCKY, Karl-Uwe (KIT); Dr SÜSS, Wolfgang (KIT); STEINMEIER, Leon (HZDR); KOUBAA, Mohamed Anis (KIT)

Presenter: Dr SÜSS, Wolfgang (KIT)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **31** Contribution code: **1-06** Type: **Poster**

ELN-DIY-Meta: Creating Interoperability for ELNs

Electronic lab notebooks (ELNs) serve as means to gather analog metadata, e.g. experimental parameters, that would otherwise be hard to digitalize. However, different systems are often used within the same research institution or community, especially when covering a long, interdisciplinary process chain. The use of different systems in the same institution - each addressing distinct requirements for discipline-specific needs - enables the availability of a broad functionality but results in challenges due to an often missing interoperability of the metadata. We are addressing this lack of interoperability for the two ELNs Herbie and Chemotion with an API-based data exchange.
A specific use-case in membrane research is treated as a starting point. As a first step, the necessary metadata for the use case were defined in both ELNs and their data fields implemented. A mapping of the corresponding data fields and the adaptation of general metadata schemes lead to a discipline-specific metadata exchange format being processed via the ELNs'APIs.
The entire process will be generalized in a guideline, motivating other ELN developers to implement interconnections for metadata transfer. The envisaged reduction of boundaries between different disciplines will enable the creation of large and coherent data sets in experimental research.

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Re-User

Please specify "other" (stakeholder)

In addition please add keywords.

ELN, Interoperability, Experimental Metadata

Primary authors: ESCHKE, Catriona (Helmholtz-Zentrum Hereon); KIRCHNER, Fabian (Hereon); SAHIM, Sayed Ahmad (Hereon); HELD, Martin (Hereon); JUNG, Nicole (KIT)

Presenter: HELD, Martin (Hereon)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **32** Contribution code: **1-36** Type: **Poster**

NFDI Matwerk - Reference Datasets

Within NFDI-MatWerk (“National Research Data Infrastructure for Material Sciences”/ “Nationale Forschungsdateninfrastruktur für Materialwissenschaften und Werkstofftechnik“), the Task Area Materials Data Infrastructure (TA-MDI) will provide tools and services to easily store, share, search, and analyze data and metadata. Such a digital materials environment will ensure data integrity, provenance, and authorship. The MatWerk consortium aims to develop specific solutions jointly with Participant Projects (PPs), which are scientific groups or institutes covering different domains, from theory and simulations to experiments. The Data Exploitation Methods group of the Karlsruhe Institute of Technology-Steinbuch Centre of Computing, as part of TA-MDI, is developing specific solutions in close collaboration with three PPs.

PP07, together with the University of Stuttgart, aims at the image-based prediction of the material properties of stochastic microstructures using high-performance solvers and machine learning. PP13, in cooperation with the University of Saarland, focuses on tomographic methods at various scales in materials research. PP18, together with the Federal Institute for Materials Research and Testing (“Bundesanstalt für Materialforschung und -prüfung”), aspires to define the criteria for materials reference datasets and usage analytics.

The requirements and goals are comparable for each PP: their research outputs, which are scientific datasets, should conform to the FAIR (Findable, Accessible, Interoperable, Reusable) principles. We aim to shape them from a data management perspective making use of the FAIR Digital Object concept, including structured metadata and storage solutions. The results will be a blueprint which will act as a reference for future datasets. Even though the collaboration is in an early stage, the initial steps already show the added value of this approach.

This research has been supported by the Federal Ministry of Education and Research (BMBF) –funding code M532701 / the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number NFDI 38/1, project no. 460247524.

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify ”other” (stakeholder)

In addition please add keywords.

FAIR,
NFDI,
Materials Science,
Metadata

Primary authors: SHAKEEL, Yusra (Karlsruhe Institute of Technology); SOYSAL, Mehmet; VITALI, Elias (Karlsruhe Institute of Technology); OST, Philipp (KIT); AVERSA, Rossella (Karlsruhe Institute of Technology)

Co-authors: CALDERÓN, Luis A. Ávila (Bundesanstalt für Materialforschung und -prüfung); ENGSTLER, Michael (Saarland University); FELL, Jonas (Saarland University); FRITZEN, Felix (University of Stuttgart); HERRMANN, Hans-Georg (Saarland University and Fraunhofer IZFP); LAADHAR, Amir (University of Stuttgart); OLBRICHT, Jürgen (Bundesanstalt für Materialforschung und -prüfung); PAULY, Christoph (Saarland University); ROLAND, Michael (Saarland University); SKROTZKI, Birgit (Bundesanstalt für Materialforschung und -prüfung)

Presenter: AVERSA, Rossella (Karlsruhe Institute of Technology)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **33** Contribution code: **2-41** Type: **Poster**

DAPHNE4NFDI

Recording data with the help of photons and neutrons is limited to bigger institutes. Besides the limited time slots, this process is also quite expensive. To save resources, DAPHNE4NFDI focuses on creating ontologies and infrastructure to make all data from its participants FAIR. This enables users not only to use existing data but also to automatically fetch data for analysis. This analysis process can also be started in the institute context. This way, analysis can be made repeatable as well, because the used software is stored and versioned at the institutes.
Three big building blocks for this project are ontologies, metadata catalogs and a common search across all institutes. Onthologies are used to have the same names for the same variable or technique. Meta-data catalogs essentially are databases that store meta-data of the collected data. This meta-data describes the environment the data was collected in. The common search, along with the ontologies and meta-data catalogs, then enable users around the world to search for data by its meta-data.
A problem that other projects share is the sharing and tracking of data across multiple instances. If a sample is created in one institute and then taken to another one, the data has to be shared. Should the sample be altered or destroyed, this change has to be communicated to everyone else in order to save the whole life-cycle.
At JCNS we use the instrument control software Nicos. Nicos implements the concepts of datasinks, which enables us to save the recorded data in multiple ways. To decouple Nicos and the meta-data catalog, we plan to use a structure that buffers every request. The two biggest advantages are network connection independence and logging of all operations.

Please assign your poster to one of the following keywords.

Standards

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify ”other” (stakeholder)

In addition please add keywords.

NFDI

Primary author: HANNEMANN, Moritz Valentin (Forschungszentrum Jülich)

Presenter: HANNEMANN, Moritz Valentin (Forschungszentrum Jülich)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **34** Contribution code: **1-24** Type: **Poster**

FAIR Data Commons / Essential Services and Tools for Metadata Management Supporting Science

A sophisticated ensemble of services and tools enables high-level research data and research metadata management in science. On a technical level, research datasets need to be registered, preserved, and made interactively accessible using repositories that meet the specific requirements of scientists in terms of flexibility and performance. These requirements are fulfilled by the Base Repo and the MetaStore of the KIT Data Manager Framework.
In our data management architecture, data and metadata are represented as FAIR Digital Objects that are machine actionable. The Typed PID Maker and the FAIR Digital Object Lab provide support for the creation and management of data objects. Other tools enable editing of metadata documents, annotation of data and metadata, building collections of data objects, and creating controlled vocabularies.
Information systems such as the Metadata Standards Catalog and the Data Collections Explorer help researchers select domain-specific metadata standards and schemas and identify data collections of interest.
Infrastructure developers search the Catalog of Repository Systems for information on modern repository systems, and the FAIR Digital Object Cookbook for recipes for creating FAIR Digital Objects.
Existing knowledge about metadata management, services, tools, and information systems has been applied to create research data management architectures for a variety of fields, including digital humanities, materials science, biology, and nanoscience. For Scanning Electron Microscopy, Transmission Electron Microscopy and Magnetic Resonance Imaging, metadata schemas were developed in close cooperation with the domain specialists and incorporated in the research data management architectures.
This research has been supported by the research program ‘Engineering Digital Futures’of the Helmholtz Association of German Research Centers, the Helmholtz Metadata Collaboration (HMC) Platform, the German National Research Data Infrastructure (NFDI), the German Research Foundation (DFG) and the Joint Lab “Integrated Model and Data Driven Materials Characterization (MDMC)”. Also, this project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 101007417 within the framework of the NFFA-Europe Pilot (NEP) Joint Activities.

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify ”other” (stakeholder)

data producer, data user

In addition please add keywords.

FAIR, services, tools, metadata

Primary author: STOTZKA, Rainer (KIT)

Co-authors: PFEIL, Andreas (KIT); TONNE, Danah (KIT); VITALI, Elias (KIT); ERNST, Felix (KIT); GÖTZELMANN, Germaine (KIT); ABDILDINA, Gulzaure (KIT); FRANK, Laura (KIT); DUDA, Leonhard (KIT); INCKMANN, Maximilian (KIT); SOYSAL, Mehmet (KIT); BLUMENRÖHR, Nicolas (KIT); OST, Philipp (KIT); TÖGEL, Philipp (KIT); JOSEPH, Reetu (KIT); AVERSA, Rossella; CHELBI, Sabrine (KIT); JEJKAL, Thomas; JHA, Vandana; HARTMANN, Volker (KIT); SHAKEEL, Yusra (KIT)

Presenter: STOTZKA, Rainer (KIT)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **35** Contribution code: **1-25** Type: **Poster**

NFDI MatWerk / Materials Data Infrastructure

The German National Research Data Infrastructure (NFDI) aims to systematically develop sustainably secure and make accessible the data holdings of science and research. It is being established as a networked structure of consortia acting on their own initiative. In NFDI-MatWerk, a reliable digital platform for the materials and nanosciences is being established, which enables the digital representation of materials data and specific metadata. Within NFDI-MatWerk the Task Area Materials Data Infrastructure will provide services to easily store, share, search, and analyze data and metadata while ensuring data integrity, provenance, and authorship.

The concept of FAIR Digital Objects, developed in the Research Data Alliance and in the FAIR Data Commons of HMC, will be utilized to represent data objects. Data sets and metadata documents will be stored in research data repositories and metadata repositories, respectively.
Metadata is one of the key elements to implement both human-readable as well as machine-actionable representations of materials-related information. Additional services will be provided for metadata enrichment and annotation, harvesting and indexing, as well as for documenting the provenance of the data objects. Collections of FAIR Digital Objects will be fed into a knowledge graph based on relevant Materials Science and Engineering ontologies connecting materials information and data.
Web front-ends will provide access to data, optimized for the particular perspectives of the user groups. Support and training will be provided for the use as well as the operation of the Materials Data Infrastructure services and tools.
First adopters of the research data and metadata infrastructures are participant projects providing data sets from various fields that will be transformed into exemplary reference data sets.

This research has been supported by the research program 'Engineering Digital Futures'of the Helmholtz Association of German Research Centers, the Helmholtz Metadata Collaboration (HMC) Platform, the German National Research Data Infrastructure (NFDI), and the German Research Foundation (DFG).

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify "other" (stakeholder)

In addition please add keywords.

FAIR, NFDI, Metadata, Materials Science

Primary author: STOTZKA, Rainer

Co-authors: MOGHADDAM, Amirreza (RWTH); VITALI, Elias (KIT); BITZEK, Erik (MPIE/FAU); GRÜNWALD, Katharina (RWTH); POLITZE, Marius (RWTH); SOYSAL, Mehmet; GOLOWIN, Nadine (KIT); OST, Philipp (KIT); JOSEPH, Reetu (KIT); AVERSA, Rossella; HUNKE, Sirieam (RWTH); SHAKEEL, Yusra (KIT)

Presenter: STOTZKA, Rainer

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **36** Contribution code: **1-23** Type: **Poster**

Data Collections Explorer

For research data to be used efficiently, it must be easy to find and access. This is a requirement in all areas of science. The Data Collections Explorer, developed within NFDI4Ing for the engineering sciences, targets these needs. It is an information system that provides an overview of research data repositories, archives, databases as well as individual datasets published in the field. Two use cases are considered:
1. Scientists searching for data sets. Are there datasets available to aid in your research? Are there benchmarks available to check your results? Are these datasets available under an open access license?
2. Scientists aiming to publish data sets: Among community-specific repositories, which ones are suitable to publish the research data? Do repositories restrict the size of the datasets that can be uploaded and if so, what are the limits? Are publication fees charged and if so, how much is charged?
To facilitate answering these questions, the Data Collections Explorer provides both a free text search and filters for type of service, subject area, and access license. Where appropriate and available, information on data size limits and publishing fees is provided.
The Data Collections Explorer complements re3data, as it includes entries which are outside its scope or which are not listed.

This concept is not limited to the engineering sciences. To broaden the impact, we are currently working on expanding the Data Collections Explorer to the material sciences and engineering community within NFDI-MatWerk.

This work is supported by the NFDI4Ing consortium, the German Research Foundation (DFG), the research program 'Engineering Digital Futures'by the Helmholtz Research Association and the Helmholtz Metadata Collaboration (HMC) Platform.

Please specify "other" (stakeholder)

In addition please add keywords.

NFDI, NFDI4Ing, FAIR, Engineering Sciences

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Primary author: OST, Philipp-Joachim (Karlsruhe Institute of Technology)

Presenter: OST, Philipp-Joachim (Karlsruhe Institute of Technology)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **37** Contribution code: **1-07** Type: **Poster**

riaf — a Repository Infrastructure that Accommodates Files

riaf is a repository infrastructure to accommodate files. It enables to hold the data with the FAIR principles (see also fair-principles).

riaf is designed to enable provenance and reproducibility of the research data in the early part of the data life cycle, i. e. prior to publication. It further is designed to enable checks on metadata relevant to research data management as defined e.g. in a machine actionable data management plan (maDMP).

This new concept of using CI pipelines for research data allows interesting features. The server could do cryptographic timestamping to inhibit silent changes of the history. Research data management can define relevant checks on metadata. From given metadata a public accessible landing page can be created.

In our concept most data is stored in a repository and can be easily distributed. This allows the data genesis in a private environment (e. g. aircraft, campaigns, …) without network access and share later the data using a central server instance. Also already during data genesis (e. g. raw data, physical data, scientific data) the possibility to share data and track changes is given. And in the end after preparing a publication the data can be transported to a public data repository.

The primary focus is to work as an in-house solution to handle digital assets. It should be possible to use the data without downloading a complete digital asset.

For this purpose we use open source software in a composability design (Unix philosophy):

- gitolite
- fuse_git_bare_fs
- sskm
- gitolite_web_interface
- pydabu
- git
- WebDAV
- git-annex
- OpenSSH
- apache

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Producer

Please specify ”other” (stakeholder)

In addition please add keywords.

repository data fair CI pipeline

Primary author: MOHR, Daniel (Deutsches Zentrum für Luft- und Raumfahrt e. V.)

Co-author: BRÖTZ, Björn (Deutsches Zentrum für Luft- und Raumfahrt e. V.)

Presenter: MOHR, Daniel (Deutsches Zentrum für Luft- und Raumfahrt e. V.)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **38** Contribution code: **1-34** Type: **Poster**

inst.dlr: a semantic instrument database for scientific large-scale facilities

In the data-intensive engineering and natural sciences, the precise and ontologically most comprehensive description of data through metadata is indispensable. Over the entire data life cycle, it is only complete and reliable metadata that ensure comprehensive use and re-use.
In practice, however, the gap between fundamental considerations of data management and the challenge with incomplete descriptions of often changing experiments and plants becomes apparent. The challenge and extra work of recording parameters and properties manually, often retrospectively in analogue or digital notes, where they become untraceable or incomprehensible at the latest when the employees leave.
The present project therefore starts exactly at the immediate beginning of data genesis - the measuring instrument, sensor, detector and the associated device. Even before the electronic laboratory notebook, it should help researchers to ensure that as much as possible, as reliably as possible, is already pre-recorded and pre-entered within the ELN.
inst.dlr is both a source and a repository for metadata of an instrument and its semantic dependencies and construction. It aims to support researchers in searching, selecting and documenting individual instruments, and thus in responding to specific system setups.

The three-year project will use pilot facilities to explore the challenges of data exchange, data protection and interoperability in particular. Concepts, tools, workflows and integrated advice and training will be developed for researchers and put to use.
Another focus will be on the reuse of the data provided in inst.dlr through applications in technology marketing, investment management and research transfer.

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify ”other” (stakeholder)

In addition please add keywords.

PIDINST, instrument ontology, research transfer

Primary authors: ARNDT, Witold (DLR); Dr LANGENBACH, Christian (DLR)

Presenter: ARNDT, Witold (DLR)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **39** Contribution code: **1-26** Type: **Poster**

New Approaches to Scalable FAIRification of Sample Data

Physical samples with informative metadata are more easily discoverable, shareable, and reusable. Metadata provides the framework for consistent, systematic, and standardized collection and documentation of sample information. This poster explores practical implementation of the FAIR Principles through creation of a framework centralized around biospecimens, linked datasets, sample information, and PIDs (Persistent Identifiers).

Two initiatives aimed at enhancing FAIRification of sample data will be described. The first, by SciLifeLab Data Centre, is to mobilize the community to identify a minimum set of attributes required for describing biospecimens with ontological mapping for semantic unambiguity and machine actionability. The goal is to facilitate interoperability and portability of sample information among multiple repositories and resources (e.g., e-Infrastructures). In addition, identifying the required attributes for registering biospecimen PIDs will enable coupling of descriptive metadata and objects in a FAIR and comprehensive manner.

The second initiative is development of the Inventory module of the RSpace electronic research notebook, which enables user-friendly and scalable sample collection and management and association of sample data and metadata with experimental data. The goal is similar to SciLifeLab's: to facilitate portability and interoperability of sample information—in this case between RSpace and other tools, repositories, and e-Infrastructures.

Ultimately, both initiatives implement common elements to FAIRify sample data:

- Association of variable domain-specific PIDs with sample data.
- Incorporation of variable but standardized sample metadata formats that enable scalable submission to domain repositories.
- User-friendly collection of sample data in the field.
- Automated and scalable passage of sample data and metadata through systems and into external tools and resources.

Challenges faced and approaches to overcoming them will be outlined. The role of the IGSN–DataCite partnership, which is supporting global adoption, implementation, and use of IGSN identifiers by ensuring ongoing sustainability of IGSN PID infrastructure and by fostering a 'Community of Communities of Practice'across research domains, will be highlighted.

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify "other" (stakeholder)

In addition please add keywords.

PIDS, Controlled vocabularies, Samples, Metadata

Primary authors: MACNEIL, Rory (Research Space); Mr EDMUNDS, Rorie (DataCite); Dr EL-GEBALI, Sara (SciLifeLab)

Presenters: MACNEIL, Rory (Research Space); Mr EDMUNDS, Rorie (DataCite); Dr EL-GEBALI, Sara (SciLifeLab)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **40** Contribution code: **2-36** Type: **Poster**

FAIR-DOscope –Explore the facets of FAIR Digital Objects

Working in the realm of FAIR Digital Objects can be very abstract and sometimes overwhelming. There are so many aspects which have to be addressed in order to create a first FAIR Digital Object. And if this is done, the only thing you get is a PID expected to be machine actionable by metadata available in its PID record. But which metadata is in there? Were all fields properly filled? Which relationships exist with other FAIR Digital Objects? For someone who is taking his/her first steps towards FAIR Digital Objects, answering these questions in an easy and user-friendly way is crucial to fulfill his/her task and FAIR-DOscope is supposed to give answers to these questions.
With FAIR-DOscope we provide a tool which resolves and visualizes FAIR Digital Objects in any Web browser. The result is rendered in two different ways: a tabular view showing the content of the PID record and a graphical view showing links to related FAIR Digital Objects. Both views allow further interaction, e.g., redirect to data type information stored in a Data Type registry or navigating through the FAIR Digital Object graph by clicking its nodes.
FAIR-DOscope may be used locally or provided as a remotely accessible service via the internet.
In the poster we will show how FAIR-DOscope looks like, how it can be used and what happens in the background to obtain the information presented to the user. Furthermore, we will give an outlook on our future plans for this tool, which already offers a very easy entry point for consuming FAIR Digital Objects and will be the basis for opening them also to scientists of the Helmholtz Association and beyond.
This work has been supported by the research program 'Engineering Digital Futures' of the Helmholtz Association of German Research Centers and the Helmholtz Metadata Collaboration Platform.

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify "other" (stakeholder)

Scientist/Data Service Developer

In addition please add keywords.

FAIR Digital Objects User Interface

Primary author: JEJKAL, Thomas

Presenter: JEJKAL, Thomas

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **41** Contribution code: **2-37** Type: **Poster**

The Helmholtz Kernel Information Profile - FAIR Digital Objects for the Helmholtz Association

In the concept of FAIR Digital Objects, PID Kernel Information is key to machine actionability of digital content. The PID Kernel Information is directly stored in the PID record in the database of the PID resolution service. One of the most important properties is the Data Type that allows PID Kernel Information to be used by machines for fast decision-making. To make a first step into the direction of standardizing PID Kernel Information, the RDA Working Group on PID Kernel Information has defined a first proposal of a core Kernel Information Profile (KIP). Among other aspects, the group defined a list of seven guiding principles, helping to decide on which information could be part of a KIP and which information should be stored elsewhere.
In order to reach the goals of HMC, to make the depth and breadth of research data produced by Helmholtz Centres findable, accessible, interoperable, and reusable (FAIR) for the whole science community, a common Helmholtz KIP has been agreed on serving as basis for all FAIR Digital Objects created in the context of HMC. This poster describes the Helmholtz KIP and elaborates on decisions leading to differences compared to the core KIP recommended by the RDA. While remaining mostly compatible with the RDA core KIP, the Helmholtz KIP adds some additional properties that satisfy the multidisciplinary research fields of the Helmholtz Association. Thus, it serves as a good starting point for rolling out the FAIR Digital Object concept over all Research Data Management Infrastructures of the Helmholtz Association and beyond.
In addition, the poster provides a first impression of a demonstrator, which is currently under development and should serve as a showcase.
This work has been supported by the research program 'Engineering Digital Futures' of the Helmholtz Association of German Research Centers and the Helmholtz Metadata Collaboration Platform.

Please assign your poster to one of the following keywords.

Semantics

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify "other" (stakeholder)

Scientist/Data Services Developer

In addition please add keywords.

FAIR Digital Objects PID

Primary authors: PFEIL, Andreas (KIT); JEJKAL, Thomas; PIROGOV, Anton (Forschungszentrum Jülich); Mr KOCH, Christian (Deutsches Krebsforschungszentrum); CURDT, Constanze (Helmholtz

Metadata Collaboration (HMC), GEOMAR Helmholtz Centre for Ocean Research Kiel); KREBS, Florian (DLR e.V.); Mr GÜNTHER, Gerrit (Helmholtz-Zentrum Berlin für Materialien und Energie); SCHWEIKERT, Jan (KIT); Mr WEINELT, Martin (GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel); VIDEGAIN BARRANCO, Pedro (Forschungszentrum Jülich)

Presenter: JEJKAL, Thomas

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **42** Contribution code: **1-35** Type: **Poster**

Open, Metadata-Enriched, Non-Proprietary Data Format for Data Dissemination

Researchers in the social sciences use various software for statistical analysis of rectangular, structured data . The various data formats which are only partially compatible impede data exchange and reuse. In particular, proprietary data formats endanger those in the FAIR principles enshrined demand for interoperability. The project Open Data Format aims to develop a non-proprietary Open Data Format enriched with multi-level metadata that can be collectively used in popular statistical software. At the same time, The Open Data Format can be enriched with multilingual metadata as well as further links to data portals, which allows direct access to online documentation materials via the statistical software itself. The project is in the frame of KonsortSWD, consortium for the social, behavioural, educational and economic sciences and is funded by NFDI. In this conference we would like to introduce our work from current stage, including specification of metadata profile and development of technical import and export filters for the statistical program R and Stata.

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify ”other” (stakeholder)

In addition please add keywords.

Open Data, Metadata, FAIR

Primary authors: Ms SAALBACH, Claudia (DIW Berlin); HAN, Xiaoyao (DIW Berlin)

Presenters: Ms SAALBACH, Claudia (DIW Berlin); HAN, Xiaoyao (DIW Berlin)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **43** Contribution code: **1-27** Type: **Poster**

PIDA: A lightweight PID service for sustainable findability of digital assets on the web

Web addresses and the stability of established links on the web can decay over time, for example, if a digital resource is moved to another location or when the domain name of an organization changes. This poses a large findability issue: in the case of the above, digital assets, even though links were previously well established, would effectively not be findable anymore. This problem can be mitigated by using persistent identifiers (PIDs). Instead of pointing directly to the location of an internet resource, a PID points to an intermediate resolution service. The resolution service associates the PID with the actual URL of the resource and returns that location to the client. The client can then complete the transaction. When the web address of a digital resource changes, this can be easily updated in the resolution service, all PIDs remain intact and functional, and the resource remains accessible.
In this work, we present PIDA (Persistent Identifiers for Digital Assets), a lightweight service by HMC providing unique persistent URLs (PURLs) for referencing digital assets on the web, including articles, datasets, videos, persons, or organizations. Using PURLs from PIDA ensures that digital assets remain findable and can be accessed reliably by both humans and machines in the long term. PIDA provides content negotiation for use in semantic web contexts (e.g. for ontology development). HMC is committed to keeping operating and maintaining PIDA as one of its services for the upcoming 10+ years. Get your PIDA PURL at https://purls.helmholtz-metadaten.de!

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify ”other” (stakeholder)

HMC Staff

In addition please add keywords.

persistent identifiers, findability, PURLs,content negotiation

Primary authors: FATHALLA, Said; Prof. SANDFELD, Stefan (Instititute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany.); HOFMANN, Volker

Presenter: FATHALLA, Said

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **44** Contribution code: **1-04** Type: **Poster**

Metadata in the Research Workflow: Tools for Enrichment and Validation of Structured Metadata

Improving research data management practices is both an organizational and a technical challenge: even in the same research field, (meta)data is often created, stored and processed in an ad-hoc manner. This results in a lack of a clear structure and standardization and makes the metadata "unFAIR" . We present two tools that assist scientists in their research workflows to enrich, structure and validate their data and metadata. This increases machine interpretability and reusability, e.g. to ease (automatic) data analysis or metadata harvesting pipelines.

Metador is a web-based structured submission interface for uploading research data and linking it to predefined metadata in a structured form. Metadata is supplied by completing a form for each uploaded file. The form is configurable by JSON Schemas and can adjusted by the user depending on the type of the uploaded file. This ensures that captured metadata is specific to the uploaded file type and appropriate to the scientific domain. It is intended for deployment in research groups and designed for quick and easy integration into existing scientific workflows.
Currently we are extending Metador architecture and functionality into a versatile RDM platform focused on metadata standardization and harmonization. It will be designed as an open and extensible ecosystem of reusable generic building blocks and ready-to-deploy services. Combined, they will cover aspects from initial collection of metadata up to improved search, data extraction and data visualization.

DirSchema is a specification and validation tool that enforces requirements concerning the directory structure and metadata provided in datasets. It is intended to be used by researchers and research groups during dataset generation or preparation to harmonize metadata in datasets across users and groups. DirSchema can be used by individual researchers or research groups to validate their dataset directory structures against an agreed-upon JSON Schema based specification that is provided as a YAML file. Further it can be deployed as a building block in other local or web-based scenarios to perform the validation automatically.

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify "other" (stakeholder)

HMC Staff (FAIR Data Commons)

In addition please add keywords.

metadata, validation, JSON Schema, YAML

Primary authors: PIROGOV, Anton (Forschungszentrum Jülich GmbH); Ms D'MELLO, Fiona (Forschungszentrum Jülich GmbH); Mr SANDFELD, Stefan (Forschungszentrum Jülich GmbH); Mr HOFMANN, Volker (Forschungszentrum Jülich GmbH)

Presenter: PIROGOV, Anton (Forschungszentrum Jülich GmbH)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **45** Contribution code: **2-43** Type: **Poster**

Fundamentals of scientific metadata - an entry-level training course for early career researchers

Get your hands dirty with semi-structured metadata in HMC's remote training course "Fundamentals of scientific metadata: why context matters"!

Have you ever struggled to make sense of research data provided by a collaborator - or even to make sense of your own data 5 months after publication? Do you see difficulties in meeting data description requirements of your funding agency? Do you want your data to have lasting value, but don't know how to ensure that with metadata? Our training course is here to help!

In our course, you will learn about how metadata will play a critical role in tomorrow's research and engage in hands-on tasks that introduce basic concepts related to machine-readable metadata, including:

- differences between data & metadata
- annotation of research data with metadata
- finding and evaluating suitable metadata frameworks and data repositories
- using basic Markdown / JSON / XML
- application of suitable tools for metadata annotation
- importance of semi-structured metadata and its benefits for overall scientific visibility

Get a **sneak peak into our hands-on tasks at our poster** and discover what HMC is planning for the future!

Please assign your poster to one of the following keywords.

other

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify "other" (stakeholder)

HMC staff

In addition please add keywords.

training
JSON
metadata
HMC
reproducibility

Primary authors: GERLICH, Silke (HMC); STRUPP, Annika

Co-authors: HOFMANN, Volker; SANDFELD, Stefan (Instititute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany.)

Presenters: GERLICH, Silke (HMC); STRUPP, Annika

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **46** Contribution code: **2-23** Type: **Poster**

EM Glossary: A community effort towards coordinated semantics in the electron microscopies

Semantic interoperability is one of the major challenges in implementing the FAIR principles 1 for research data. This is especially relevant for interdisciplinary projects, where people from different but related disciplines may use technical terms with differing meaning. Established vocabularies and semantic standards can harmonize domain-specific language and facilitate common understanding.
Electron microscopy (EM), as a fast-developing and widely used technique, with application across disciplines, lacks broadly applicable, formal and standardized metadata terminology. This limits research data interoperability and shows the need for terminology harmonization across the community.

Coordinated by the Helmholtz Metadata Collaboration (HMC) the EM glossary group 2 brings together researchers from more than 22 institutions across Switzerland, Austria and Germany. In a broad community effort, they work together towards a joint resource to harmonize semantics in the field of electron and ion microscopies. The EM glossary group strives for consensus on domain-specific terms via a remote, collaborative workflow on the platform GitLab. The developed resource aims to provide harmonized and machine-actionable semantics to act as a glue technology that can fundamentally support development efforts such as metadata schemas or ontologies in their respective fields.

1.Wilkinson, M.D. et al. Scientific Data. http://dx.doi.org/10.1038/sdata.2016.18 (2016)
2.EM Glossary GitLab repository: https://gitlab.hzdr.de/em_glossary

Funding statement: This work was carried out at the Hub Information and the Hub Matter of the Helmholtz Metadata Collaboration (HMC) Platform

Please assign your poster to one of the following keywords.

Semantics

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify ”other” (stakeholder)

HMC Staff

In addition please add keywords.

electron microscopy, glossary, community, interoperability

Primary authors: STRUPP, Annika (Institute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany); MANNIX, Oonagh (Helmholtz

Zentrum Berlin, Berlin Germany); HOFMANN, Volker (Institute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany)

Co-author: Prof. SANDFELD, Stefan (Institute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany)

Presenter: STRUPP, Annika (Institute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **47** Contribution code: **2-34** Type: **Poster**

FAIR DO Application Case for Composing Machine Learning Training Data

The application case for implementing and using the FAIR Digital Object (FAIR DO) concept aims to simplify usage of label information for composing Machine Learning (ML) training data.
Image data sets curated by different domain experts usually have non-identical label terms. This prevents images with similar labels from being easily assigned to the same category. Therefore, using the images collectively for application as training data in ML comes with the cost of laborious relabeling. To automate this process, machine-actionable decisions for label information must be enabled. For this purpose the FAIR DO concept is used. A FAIR DO is a representation of scientific data and requires at least a globally unique Persistent Identifier (PID), relevant metadata, and a type.
We show the requirements for specifying and using FAIR DOs when applied to ML data. Based on an application case with Scanning Electron Microscopy (SEM) image data, a Proof-of-Principle approach shows the potential of the concept for usage in ML related data management.
This work has been supported by the research program 'Engineering Digital Futures' of the Helmholtz Association of German Research Centers and the Helmholtz Metadata Collaboration (HMC) Platform.

Please assign your poster to one of the following keywords.

Processes/Policies

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Producer

Please specify "other" (stakeholder)

In addition please add keywords.

FAIR, Metadata, image data, label,

Primary author: BLUMENROEHR, Nicolas (Karlsruhe Institute of Technology, Steinbuch Centre for Computing)

Co-authors: PFEIL, Andreas (KIT); STOTZKA, Rainer; JEJKAL, Thomas

Presenter: BLUMENROEHR, Nicolas (Karlsruhe Institute of Technology, Steinbuch Centre for Computing)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **48** Contribution code: **2-35** Type: **Poster**

Using Schema-based Metadata for Image Labels accessed with FAIR Digital Objects

Scientific image data sets can be continuously enriched by labels describing new features which are relevant for some specific task. This process can be automated by means of Machine Learning (ML) techniques. Although such an approach shows clear advantages, especially when it is applied to large datasets, it also poses an important challenge:
Relabeling image data sets curated by different scientists, in order to collectively use them for ML, requires a common agreement on the labels which can be used. This can be achieved thanks to the use of a standardized way to describe the label information: a metadata schema including vocabularies. Furthermore, machine-actionable decisions on the label information for relabeling can be enabled by the representation of images and schema-based metadata as FAIR Digital Objects (DOs).
We introduce a metadata schema including vocabularies to describe ML image data represented as FAIR DOs that can be accessed for relabeling. The specifications of the metadata schema are presented. The relevance of a standardized metadata description including vocabularies for relabeling ML image data is emphasized. It is shown how the metadata is accessed with FAIR DOs and how vocabularies support automated relabeling. This contribution supplements the content of "FAIR DO Application Case for Composing Machine Learning Training Data"with a focus on the semantic aspects for relabeling.
This work has been supported by the research program 'Engineering Digital Futures'of the Helmholtz Association of German Research Centers and the Helmholtz Metadata Collaboration Platform. This project has received funding from the 'European Union's Horizon 2020'research and innovation program under grant agreement No. 101007417 within the framework of the 'NFFA-Europe Pilot '(NEP) Joint Activities.

Please assign your poster to one of the following keywords.

Processes/Policies

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Producer

Please specify "other" (stakeholder)

In addition please add keywords.

FAIR, Machine Learning, semantics

Primary authors: BLUMENROEHR, Nicolas (Karlsruhe Institute of Technology, Steinbuch Centre

for Computing); AVERSA, Rossella (Karlsruhe Institute of Technology)

Presenter: BLUMENROEHR, Nicolas (Karlsruhe Institute of Technology, Steinbuch Centre for Computing)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **49** Contribution code: **1-01** Type: **Poster**

FAIR WISH project –Developing templates to register IGSNs for various sample types

The International Generic Sample Number (IGSN) is a unique and persistent identifier for –originally –geological samples. Recently, interest has grown to make the IGSN available for more sample types from further scientific communities from the Earth and Environment (E & E). The IGSN Metadata Schema is modular: The mandatory registration schema is complemented by the IGSN Description Schema and possible additional extensions by allocating agents. While the specific GFZ sample registration schema has been extended to include cores and drilling methods, we will extend it to also register water and vegetation sample types within the HMC project "FAIR Workflows to establish IGSN for Samples in the Helmholtz Association (FAIR WISH)". Furthermore, we will use existing and controlled vocabularies, to make the registered data findable. More information on these controlled vocabularies can be found in the FAIR WISH D1 –List of identified linked open data vocabularies to be included in IGSN metadata (https://doi.org/10.5281/zenodo.6787200).
The templates to register IGSNs for samples should ideally fit to various sample types. In a first step, we created templates for samples from surface water and vegetation from AWI polar expeditions on land (AWI Use Case). We incorporated the two other FAIR WISH use cases with core material from the Ketzin coring site (Ketzin Use Case) and for a wide range of marine biogeochemical samples (Hereon Use Case) as well as further sample types from lakes and vegetation sites (AWI Use Case). These templates will be distributed for discussion and published within a wide range of E & E communities.
In our poster, we will present the templates and invite everyone to discuss the applicability to the various sample types. This discussion will go on in a dedicated user workshop in fall 2022.

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Re-User

Please specify "other" (stakeholder)

In addition please add keywords.

FAIR WISH
IGSN
Template

Primary authors: BRAUSER, Alexander (Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZen-

trum GFZ); Dr HEIM, Birgit (Alfred-Wegener-Institut Hemlholtz-Zentrum für Polar- und Meeresforschung); Dr WIECZOREK, Mareike (Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung); Dr ELGER, Kirsten (Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ); Mrs BALDEWEIN, Linda (Helmholtz-Zentrum Hereon); FRENZEL, Simone (Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ); KLEEBERG, Ulrike (Helmholtz-Zentrum Hereon)

Presenter: Dr WIECZOREK, Mareike (Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **50** Contribution code: **1-41** Type: **Poster**

Helmholtz Imaging

Helmholtz Imaging's mission is to unlock the potential of imaging in the Helmholtz Association. Image data provide a substantial part of data being generated in scientific research. Helmholtz Imaging is the overarching platform to better leverage and make accessible to everyone the innovative modalities, methodological richness, outstanding expertise and data treasures of the Helmholtz Association.

Helmholtz Imaging empowers and supports scientists in all aspects of imaging, on different occasions, at any point in their career and at all levels. Imaging-based research projects are encouraged to contact us for scientific or technical support, to enter into collaborations with our research groups, or to network with imaging experts from other Helmholtz programs.
Discover our portfolio and become a part of Helmholtz Imaging!

Please assign your poster to one of the following keywords.

other

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify "other" (stakeholder)

Helmholtz Imaging

In addition please add keywords.

Incubator, Platform, Imaging, AI, Helmholtz

Primary authors: SCHMIDT, Deborah; ISENSEE, Fabian (HIP Applied Computer Vision Lab, Division of Medical Image Computing, German Cancer Research Center); SANDER, Knut (Helmholtz Imaging); HEUSER, Philipp (Helmholtz Imaging/DESY); KRAUSE-SOLBERG, Sara (Helmholtz Imaging, DESY)

Presenter: HEUSER, Philipp (Helmholtz Imaging/DESY)

Track Classification: Postersession

Contribution ID: **52** Contribution code: **1-05** Type: **Poster**

Implementing FAIR in the Domain of Energy Systems Analysis

Within the research project LOD-GOESS (https://lod-geoss.gitub.io) and the Helmholtz Metadata Hub Energy we are developing a distributed data architecture for sharing and improved discovery of research data in the domain of energy systems analysis. A central element is the databus (https://databus.dbpedia.org) which acts as a central searchable metadata catalog. Data will be annotated on the base of the Open Energy Metadata String (https://github.com/OpenEnergyPlatform/oemetadata) and with the help of the Open Energy Ontology (https://openenergy-platform.org/ontology/). The metadata string follows the idea of the frictionless data package and can describe the data down the individual collumns and rows within the tables. We have set up a number of demonstrators which show how the infrastructure can be used to publish, search and discover data based on semantic searches, how it can be used to develop and share a common technology data base or to couple different models based on a common description of the data.

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify ”other” (stakeholder)

In addition please add keywords.

Energy Systems Analysis, Search, Discovery

Primary author: HOYER-KLICK, Carsten (German Aerospace Center, Institute of networked Energy Systems, Department of Energy Systems Analysis)

Presenter: HOYER-KLICK, Carsten (German Aerospace Center, Institute of networked Energy Systems, Department of Energy Systems Analysis)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **53** Contribution code: **2-21** Type: **Poster**

Domain level ontology design: DISO and MDMC-PROV

How can a computer understand the relations of data or objects from the real world? Ontologies are semantic artifacts that capture knowledge about their domain of interest in a machine-understandable form. The main goal of developing ontologies is to formalize concepts and their relations through which humans express meaning and to use them as a communication interface to machines. Thus, ontology development is an important step towards generating linked and FAIR data.
Within HMC we support and co-develop domain and application-level ontologies. Here we present two developments: Dislocation Ontology (DISO) and Model and Data-Driven Materials Characterization Provenance (MDMC-PROV).

DISO: An important class of materials is crystalline materials, e.g., metals and semiconductors, which nearly always contain defects, the "dislocations". This type of defect determines many important material properties, e.g., strength and ductility. Over the past years, significant effort has been put into understanding dislocation behavior across different length scales via experimental characterization techniques and simulations. However, there is still a lack of common standards to formally describe and represent disclocations. Thus, in this work we develop the dislocation ontology (DISO), which is a domain ontology that defines the concepts and relationships related to linear defects in crystalline materials. DISO is published 1 through a persistent URL following W3C best practices for publishing Linked data.

MDMC-Prov: The rapid development of science and technology in everyday large data generation does not match the data understanding. These days, understanding how experiments are performed and results are derived become more complex due to a lack of provenance documentation. Therefore, the provenance must be tracked, described, and managed over the research process. Thus, in this work, we report an application ontology that can capture provenance information in materials science experiments. The ontology is based on the MDMC glossary 2, which defines the common terms in the materials science experiments. From each term, we map to PROV-O 3. These ensure the validity, reproducibility, and reusability of the data.

1 https://purls.helmholtz-metadaten.de/diso

2 https://jl-mdmc-helmholtz.de

3 https://www.w3.org/TR/2013/NOTE-prov-primer-20130430/

Please assign your poster to one of the following keywords.

Semantics

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Re-User

Please specify "other" (stakeholder)

In addition please add keywords.

Ontologies

Primary author: IHSAN, Ahmad Zainul (Instititute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany.)

Co-authors: Dr FATHALLA, Said (Instititute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany.); AVERSA, Rossella (Karlsruhe Institute of Technology); Dr JALALI, Mehrdad (Karlsruhe Institute of Technology, Karlsruhe, Germany); Dr PANIGHEL, Mirco (CNR-IOM, Italy); OSMENAJ, Elda (CNR-IOM, Italy); Dr HOFMANN, Volker (Instititute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany.); Prof. SANDFELD, Stefan (Instititute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany.)

Presenter: IHSAN, Ahmad Zainul (Instititute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany.)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **54** Contribution code: **1-12** Type: **Poster**

Putting metadata to work in research with Marble and Beaverdam

Researchers in many fields rely on complex data from specialized instruments and large numbers of experiments. Metadata is key to efficiently document and describe data's essential attributes, and help to generate overviews of large datasets. Manually collecting and curating the extensive amounts of metadata required –some of which might be even inaccessible –is a major challenge. To support scientists in this endeavour, we develop software tools that automatically extract and consolidate metadata, which can then be put to use during data selection and further processing.

Marble (MetadAta in pRoprietary Binary fiLEs) allows researchers to identify and extract metadata from proprietary binary files. Software that operates scientific instruments often stores data in proprietary formats which are inaccessible outside of the manufacturer's ecosystem. Even if the software supports data exports, exported files may contain only a subset of the metadata originally captured, thus valuable information might be lost. Marble supports researchers in identifying sequences of data and deciphering metadata from proprietary data formats. A user interface presents results to researchers, who can adjust the deciphering method and annotate the identified information. The method and file structure is stored in converter files, which can be reused and exchanged between users. Our long-term goal is to create a software framework which creates fully automatic converters to translate proprietary data into accessible and reusable formats.

Beaverdam (Build, Explore, And Visualize ExpeRimental DAtabases of Metadata) allows researchers to interactively explore large amounts of metadata. Users can combine similar metadata from multiple experiments into a database, then access a graphical user interface in a web browser to interactively identify subsets of experiments meeting specific criteria. A data table and interactive plots describe characteristics of the selected experiments on the fly. Beaverdam is suitable for use at multiple scales –individual researchers can build and access a database locally, or a research group can maintain a joint database which members access remotely. While we are testing Beaverdam on electrophysiological brain data, it is domain agnostic and will be useful for all research disciplines.

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify ”other” (stakeholder)

HMC staff

In addition please add keywords.

metadata, software, proprietary, database, dashboard

Primary authors: D'MELLO, Fiona (Forschungszentrum Jülich GmbH); MORE, Heather (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany; Institute for Advanced Simulation (IAS-9), Jülich Research Centre, Jülich, Germany); BRINCKMANN, Steffen (Institute for Energy and Climate Research –Structure and Function of Materials (IEK-2), Jülich Research Centre, Jülich, Germany); GRÜN, Sonja (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany); DENKER, Michael (INM-10, Forschungszentrum Jülich); HOFMANN, Volker; SANDFELD, Stefan (Instititute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany.)

Presenters: D'MELLO, Fiona (Forschungszentrum Jülich GmbH); MORE, Heather (Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany; Institute for Advanced Simulation (IAS-9), Jülich Research Centre, Jülich, Germany)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **55** Contribution code: **1-38** Type: **Poster**

Quantifying FAIRness: evaluating Helmholtz data repositories using F-UJI

For research data to be reusable by scientists or machines, the data and associated meta-data should comply with the so-called "FAIR principles", i.e. it should be findable, accessible, interoperable, and reusable 1. To realize this, is not a straightforward task, as researchers do not know how FAIR or un-fair their data actually is and how to improve their FAIRness. A quantitative measure, which is easy to apply could help. The Helmholtz Metadata Collaboration (HMC) works on improving tools to automate the assessment of the FAIRness of publications.

The F-UJI framework 2 originating from the FAIRsFAIR project is a powerful tool that provides a score for the FAIRness for machine findable and readible metadata of a given publication with respect to the FAIRsFAIR metric 3. We co-develop F-UJI to explore and evaluate ways to apply it in user-sided tools.

On this poster we present a FAIR assessment through F-UJI of Helmholtz data repositories. With our work, we want to identify gaps in the Helmholtz data infrastructure with respect to the FAIRness of (meta)data and how these gaps can be closed effectively. We also provide an outlook on possible research into the development of FAIRness over time within communities.

1 Wilkinson, M.D.et al. Sci Data 3, 160018 (2016)
2 Devaraju, A., Huber, R. (2020). F-UJI, Zenodo. https://doi.org/10.5281/zenodo.4063720
3 Devaraju, A., et al. (2020). FAIRsFAIR Metrics. Zenodo.
https://doi.org/10.5281/zenodo.6461229

Please assign your poster to one of the following keywords.

other

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify "other" (stakeholder)

HMC-Staff

In addition please add keywords.

FAIRassment, FAIR principles, Metric, FUJI

Primary authors: Dr BRÖDER, Jens (Instititute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany.); VIDEGAIN BARRANCO, Pedro (Instititute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany.); Dr HOFMANN, Volker (Instititute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany.); SANDFELD, Stefan

(Instititute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany.)

Presenter: Dr BRÖDER, Jens (Instititute for Advanced Simulation –Materials Data Science and Informatics (IAS-9); Forschungszentrum Jülich, Jülich, Germany.)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **56** Contribution code: **1-03** Type: **Poster**

FAIR Digital Object Lab for your research

The FAIR Digital Object Lab is an extendable and adjustable software stack for generic FAIR Digital Object (FAIR DO) tasks. It consists of a set of interacting components with services and tools for creation, validation, discovery, curation, and more.
Preprocessing data for research, like finding, accessing, unifying or converting, takes up to 80% of research time spans. The FAIR (Findability, Accessibility, Interoperability, Reusability) principles aim to support and facilitate the reuse of data, and are therefore tackling this problem. A FAIR Digital Object (FAIR DO) capsules research data resources of all kinds (raw data, metadata, software, …) so they are following the FAIR principles.
FAIR DOs are expressive, machine-actionable pointers to research data. As such, each FAIR DO points to one research data object. Additionally, they may link to other FAIR DOs, explaining their relations.
The creation and maintenance of FAIR DOs is not trivial, as their PIDs contain typed record information. They are meant to be machine-actionable, not human-readable. Easing the creation and maintenance of FAIR DOs, as well as making FAIR DOs searchable and human-accessible, are functions of the FAIR DO Lab.
We are developing an extendable research lab for FAIR DOs, called the “FAIR DO Lab”. Its goal is to have a production-ready and configurable software stack, easing the development of FAIR-DO-aware tools and services by fostering the described use-cases. Additionally, generic tools related to FAIR research data management will be integrated. We already gained some experience by its predecessor, the FAIR DO Testbed, which was introduced at the RDA Virtual Plenary 17 Poster Session.
This research has been supported by the Helmholtz Metadata Collaboration (HMC) Platform, the German National Research Data Infrastructure (NFDI) and the German Research Foundation (DFG).

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify ”other” (stakeholder)

In addition please add keywords.

FAIR Digital Object, Services, Tools

Primary author: PFEIL, Andreas (Karlsruhe Institute of Technology (KIT))

Co-authors: CHELBI, Sabrine (KIT); JEJKAL, Thomas

Presenter: PFEIL, Andreas (Karlsruhe Institute of Technology (KIT))

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **57** Contribution code: **1-02** Type: **Poster**

Research Object Crates: Bundling Research Data and Information

Research Object Crate (RO-Crate) is an open, community driven data package specification to describe all kinds of file-based data, as well as entities outside the package. In order to do so, it uses the widespread JSON-format, representing Linked Data (JSON-LD), allowing to link to external information. This makes the format flexible and machine-readable. These packages are being referred to as (RO-)crates.
Similar to other formats, RO-Crates is based on files and folders and has a single metadata file to describe the whole package. Therefore, such packages are easy to share between different computer systems and software.
In order to create such crates, the RO-Crate community developed libraries written in different programming languages like Python, Ruby, JavaScript, and Java. With Describo, there is also a graphical user interface available.
We developed the ro-crate-java library, which allows creating, modifying and validating crates using the Java Programming Language. The focus of development was the ease of use: We aimed to make it intuitive and easy to create valid crates, without knowing the specification too well. Our implementation can be used for integration into repositories or other services or tools.
This research has been supported by the Helmholtz Metadata Collaboration (HMC) Platform, the German National Research Data Infrastructure (NFDI) and the German Research Foundation (DFG).

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify ”other” (stakeholder)

In addition please add keywords.

Research Data Package, FAIR, Metadata

Primary authors: TZOTCHEV, Nikola; SCHOLZ, Jonas; PFEIL, Andreas (Karlsruhe Institute of Technology (KIT))

Presenter: PFEIL, Andreas (Karlsruhe Institute of Technology (KIT))

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **59** Contribution code: **1-31** Type: **Poster**

FAIR DO Cookbook

The FAIR DO Cookbook helps researchers to learn about FAIR Digital Objects (FAIR DOs) practically. It is guided by existing implementations and expert guidelines from HMC.
The FAIR (Findability, Accessibility, Interoperability, Reusability) principles aim to support and facilitate the reuse of data. They require the use of structured metadata and other important aspects in research data management. But for automation, machines do not only need to read this information, but to understand and act on it. The concept of FAIR Digital Objects (FAIR DO) aims to be a common layer for all (FAIR) data to achieve this machine-actionability. It offers an intentionally limited set of typed information to enable the machine to make decisions on the data objects:

- Can I (the machine) access it? Will access need human intervention?
- What kind of object is this? Can I open and read it? In which context was it created?
- What metadata exists for this object, and can I interpret it?
- Which actions can I perform on this object?

If a machine can answer such questions, this will further increase FAIRness as it will make it easy and reliable to build services (like research graphs or search indices) and tools for a wider range of data.
As the FAIR Data Commons group, we deal with the implementation of FAIR DOs for the Helmholtz Association. Therefore, we created the FAIR DO Cookbook to collect recipes and best practices. Issues, open questions, or suggestions can be reported in its GitHub Repository.
This research has been supported by the Helmholtz Metadata Collaboration (HMC) Platform, the German National Research Data Infrastructure (NFDI) and the German Research Foundation (DFG).

Please assign your poster to one of the following keywords.

Processes/Policies

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify ”other” (stakeholder)

In addition please add keywords.

FAIR Digital Object, Recipes, Guidelines

Primary authors: JEJKAL, Thomas; PFEIL, Andreas (Karlsruhe Institute of Technology (KIT)); HARTMANN, Volker (KIT)

Presenter: JEJKAL, Thomas

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **60** Contribution code: **2-27** Type: **Poster**

Enriched metadata for hybrid data compilations with applications to cryosphere research

In geodisciplines such as the **cryosphere** sciences, a large variety of data is available in data repositories provided on platforms such as Pangaea. In addition, many computational process models exist that capture various physical, geochemical, or biological processes at a wide range of spatial and temporal scales and provide corresponding simulation data. A natural thought is to **hybridize measured and simulated data** into comprehensive data sets that complement each other and provide a joint basis for subsequent model-based interpretation. Two aspects remain challenging, namely *a)* we are lacking a **unified metadata** approach that is ready to use for hybrid data compilations comprising both measured and simulated data each with their own characteristics and natural limitations, and *b)* we are not providing these data compilations in an '**analysis-ready**' format, for instance, including uncertainties.
In this contribution, we present an example from cryosphere science, where much potential remains in a joint interpretation of several field tests and simulation studies to generate an integrated, holistic representation of the ice body. Yet, to date, this joint interpretation is often not feasible because metadata of the measurements lack **cross-repository consistency and completeness**, and simulated data are often not equipped with metadata at all. We discuss these challenges in light of FAIR, while focusing on the example of **sea ice core data**. Specifically, we introduce our in-house Ice Data Hub (IDH) as a **flexible data management** tool that aims to overcome these challenges. We use the IDH to *a)* store measurement data sets together with enriched, consistent metadata, *b)* display, add, and plot data sets through its web browser-based GUI, and *c)* directly couple simulation environments to facilitate **interdisciplinary dataflow and interoperability**. Lastly, we present an example of an 'analysis-ready'sea ice core data set that is merged from individual ice cores stored in the IDH.

Please assign your poster to one of the following keywords.

other

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Re-User

Please specify "other" (stakeholder)

In addition please add keywords.

cryosphere, hybrid, data management, interoperability

Primary author: SIMSON, Anna (RWTH Aachen, Methods for Model-based Development in Com-

putational Engineering)

Co-authors: Prof. KOWALSKI, Julia (RWTH Aachen, Methods for Model-based Development in Computational Engineering); Dr BOXBERG, Marc S. (RWTH Aachen, Methods for Model-based Development in Computational Engineering)

Presenter: SIMSON, Anna (RWTH Aachen, Methods for Model-based Development in Computational Engineering)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **61** Contribution code: **2-08** Type: **Poster**

The GHGA Metadata Model: A Framework for National Data Sharing

Data sharing at both the national and international level benefits genomic medicine. Specifically, in diseases driven by genomic factors, such as cancer types or rare diseases, data sharing maximizes the utility and impact of cohorts, thereby aiding in translating research findings to therapies. The successful discovery of new findings, however, requires linking genomic data to health data and sharing both, which in turn necessitates national metadata standardization and harmonization along with a data protection framework.

The necessary infrastructure for FAIR (findable, accessible, interoperable, and reusable) data management, storage, and access on a national level will be provided by the German Human Genome-Phenome Archive (GHGA). In the work presented here, we introduce the underlying metadata model of GHGA. By exploring and building on several already existing models and in close discussions with stakeholders from genomic medicine, we defined a harmonized metadata model covering metadata elements pertaining to technical (experiment and analysis), individual (sample) and dataset data. Standardization of the model is achieved with the usage of several well-established ontologies and the definition of controlled vocabularies, making it self-describing, unambiguous, flexible, and expressive. The schematic model backbone is defined as YAML using the Linked Data Modeling Language (LinkML). As GHGA will be part of the federated European Genome Archive (EGA), our model is designed to be compatible with EGA.

Our model demonstrates how genomic and health data can be stored in accordance with General Data Protection Regulations and securely shared across German institutions. It balances the individual data subject's right to privacy while ensuring high-quality metadata to make genomic data in GHGA findable and reusable.

Please assign your poster to one of the following keywords.

Standards

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify "other" (stakeholder)

In addition please add keywords.

model, standardization, harmonization, ontologies, health

Primary author: Ms MAUER, Karoline (DZNE)

Co-authors: IYAPPAN, Anandhi; UNNI, Deepak; KRAUS, Florian; PARKER, Simon; TREMPER,

Galina; SURUN, Bilge; MENGES, Paul; KIRLI, Koray; SCHULTZE, Joachim. L.; ULAS, Thomas; NAHNSEN, Sven

Presenter: Ms MAUER, Karoline (DZNE)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **62** Contribution code: **2-42** Type: **Poster**

Data management practices among Helmholtz's research communities - A survey on the status quo and on community-specific demands

In autumn 2021, the Helmholtz Metadata Collaboration (HMC) concluded its first HMC Community Survey to get in touch with Helmholtz's research communities. The survey aimed at characterizing the community-specific research data management and data publication practices as well as related gaps and needs expressed by Helmholtz's research communities. For this purpose, we developed a question catalog with a total of 49 (sub-)questions, targeted at researchers with various levels of expertise. We distributed the survey among researchers in all six Helmholtz research fields and obtained 631 survey replies from 18 research centers. 1
In this poster, we will share some of the highlights of this survey data. This will include a description of the status quo of data handling, data publication and metadata practices in the six different research fields. We will further illustrate the gaps and needs as articulated by the research community in Helmholtz. HMC will continue to use the obtained data from the community survey to strategically develop its service portfolio and communication strategy towards the scientific staff and centres'requirements in Helmholtz.

1 https://doi.org/10.7802/2433

Please assign your poster to one of the following keywords.

other

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

In addition please add keywords.

Community Survey, Data Management Practices

Please specify "other" (stakeholder)

Helmholtz Metadata Collaboration, Task Force Survey

Primary authors: LEMSTER, Christine (Geomar); KULLA, Lucas (DKFZ); KUBIN, Markus (HMC, HZB); GERLICH, Silke (HMC); SÖDING, Emanuel (GEOMAR); SCHWEIKERT, Jan (KIT); STEINMEIER, Leon (Helmholtz Institute Freiberg); SHANKAR, Sangeetha (German Aerospace Center)

Presenters: KULLA, Lucas (DKFZ); GERLICH, Silke (HMC)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **63** Contribution code: **2-17** Type: **Poster**

Understanding Data Management Practices in Research Field Matter - Conclusions from a Multi-Method Approach

Supporting Helmholtz's research communities in making their data FAIR is one of the key missions of HMC. A multi-method approach combining quantitative and qualitative methods was developed to understand current data management practices in research field Matter. Quantitative information was obtained from data that was self-reported by Helmholtz's researchers in the HMC Community Survey 2021. Complementary data on Open and FAIR data practices in research field Matter, gathered in a data mining approach, is visualized in a dashboard. Qualitative understanding of community-specific FAIR data practices was obtained from a manual FAIR assessment based on the FAIR Data Maturity Model. Here we report on a combined interpretation of HMC Hub Matter's findings from this multi-method approach. Three key areas for future action by HMC Hub Matter are discussed, such as (1) bridging policy and practicability, (2) creating a culture of data reuse, and (3) monitoring and engaging with technical infrastructure.

Please assign your poster to one of the following keywords.

other

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify "other" (stakeholder)

HMC Hub Matter

In addition please add keywords.

Matter, FAIR, Data-Management, Survey, Data-Mining

Primary authors: KUBIN, Markus (HMC, HZB); GUENTHER, Gerrit (Helmholtz-Zentrum Berlin); GILEIN, Astrid; PREUSS, Gabriel (Helmholtz-Zentrum Berlin für Materialien und Energie); CRISTIANO, Luigia (HZB); Mr WALTER, Konstantin Pascal (Helmholtz-Zentrum Berlin für Materialien und Energie); SERVE, Vivien (Helmholtz-Zentrum Berlin für Materialien und Energie); GÖRZIG, Heike (HZB); MANNIX, Oonagh (HMC matter/HZB)

Presenter: KUBIN, Markus (HMC, HZB)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **64** Contribution code: **2-38** Type: **Poster**

Lessons learned from applying the FAIR Data Maturity Model to a prototypical data pipeline in Matter

A central mission of HMC is to support the data producers of the Helmholtz community in making their data FAIR. Developing a sustainable strategy for doing so requires a detailed understanding of community-specific practices, strengths, and limitations related to the application of each FAIR data guideline. We have applied the FAIR Data Maturity Model, developed by the respective RDA working group, to a prototypical data pipeline in the research field Matter. In our poster presentation, we would like to provide an overview of our approach and discuss the lessons learned that helped us identify key activities for meeting community needs.

Please assign your poster to one of the following keywords.

other

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify ”other” (stakeholder)

HMC Hub Matter

In addition please add keywords.

FAIR Assessment, RDA, FDMM, Matter

Primary authors: KUBIN, Markus (HMC, HZB); GUENTHER, Gerrit (Helmholtz-Zentrum Berlin); GÖRZIG, Heike (HZB); CRISTIANO, Luigia (HZB); MANNIX, Oonagh (HMC matter/HZB); KRAHL, Rolf (Helmholtz-Zentrum Berlin für Materialien und Energie)

Presenters: KUBIN, Markus (HMC, HZB); GUENTHER, Gerrit (Helmholtz-Zentrum Berlin)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **65** Contribution code: **2-02** Type: **Poster**

FAIR Digital Objects for 5D imagery of our and other planet(s)

Imaging the environment is an essential and crucial component in spatial science. This concerns nearly everything between the exploration of the ocean floor and investigating planetary surfaces. In and between both domains, this is applied at various scales –from microscopy through ambient imaging to remote sensing –and provides rich information for science. Due to recent the increasing number data acquisition technologies, advances in imaging capabilities, and number of platforms that provide imagery and related research data, data volume in nature science, and thus also for ocean and planetary research, is further increasing at an exponential rate. Although many datasets have already been collected and analyzed, the systematic, comparable, and transferable description of research data through metadata is still a big challenge in and for both fields. However, these descriptive elements are crucial, to enable efficient (re)use of valuable research data, prepare the scientific domains e.g. for data analytical tasks such as machine learning, big data analytics, but also to improve interdisciplinary science by other research groups not involved directly with the data collection. In order to achieve more effectiveness and efficiency in managing, interpreting, reusing and publishing imaging data, we here present a project to develop interoperable metadata recommendations in the form of FAIR 1 digital objects (FDOs) 2 for 5D (i.e. x, y, z, time, spatial reference) imagery of Earth and other planet(s). An FDO is a human and machine-readable file format for an entire image set, although it does not contain the actual image data, only references to it through persistent identifiers (FAIR marine images 3). In addition to these core metadata, further descriptive elements are required to describe and quantify the semantic content of imaging research data. Such semantic components are similarly domain-specific but again synergies are expected between Earth and planetary research. We here present the current status of the project, with the specific tasks on joint metadata description of planetary and oceanic data.

1 Wilkinson, M. I. etal. The FAIR Guiding Principles for scientific
data management and stewardship. Sci Data 3, 160018 (2016), doi:10.1038/sdata.2016.18.
2 https://fairdigitalobjectframework.org/
3 https://marine-imaging.com/fair/ifdos/iFDO-overview/

Please assign your poster to one of the following keywords.

Standards

Please specify ”other” (stakeholder)

Scientist/Data Producer (Data Infrastructure Provider)

In addition please add keywords.

Metadata, Digital Objects, GIS

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Producer

Primary authors: NASS, Andrea (DLR); SCHOENING, Timm (GEOMAR); D'AMORE, Mario (DLR); KWASNITSCHKA, Tom (GEOMAR); HAUBER, Ernst (DLR); ROATSCH, Thomas (DLR); PURSER, Autun

Presenter: NASS, Andrea (DLR)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **66** Contribution code: **1-32** Type: **Poster**

MetaStore –Managing Metadata for Digital Objects

MetaStore is a metadata repository for managing metadata documents. It supports communities in storing metadata documents in a predefined schema. It is therefore an important building block for more precise automated evaluation and/or retrieval of digital objects.
With the help of the metadata documents, digital objects can also be evaluated/compared according to content-related aspects. XML and JSON are very common as data formats for such machine-interpretable documents. However, they are only meaningful if they adhere to a certain structure and are correctly filled in. MetaStore supports the use of XML and JSON schema as the definition for the document structure. It allows you to register your own and/or existing schemas in these two formats to ensure that the documents have the appropriate structure. When ingesting metadata documents, the structure is checked and invalid documents are rejected. All valid documents are assigned a persistent identifier and can be automatically indexed for search. Public documents can be harvested via a standardized protocol (OAI-PMH).
The supplied web interface also provides a low-threshold entry point for managing documents and also allows documents to be created/edited without additional tools.
This work has been supported by the research program 'Engineering Digital Futures' of the Helmholtz Association of German Research Centers and the Helmholtz Metadata Collaboration Platform.

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify "other" (stakeholder)

In addition please add keywords.

FAIR, services, tools, metadata

Primary author: Mr HARTMANN, Volker (KIT)

Co-authors: JEJKAL, Thomas; Mrs CHELBI, Sabrine (KIT)

Presenter: Mr HARTMANN, Volker (KIT)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **67** Contribution code: **1-33** Type: **Poster**

Metadata Hub - One for all

The Metadata Hub provides a generic service for metadata repositories. Based on this, different kinds of metadata repositories can be accessed with uniform tools without the researchers having to deal with the complex details.
In the domain of research data management, there are a variety of repositories that offer metadata management services to researchers. This poses the challenge that these repositories usually have different interfaces and in nature are not very interoperable with each other, violating one of the FAIR Principles.
Our work aims to build a bridge between these repositories as a generic service for metadata repositories, the Metadata Hub. It's accessible via the Turntable API, which defines a uniform interface for metadata repositories.
To validate metadata documents, a definition of the document structure has to be available. For JSON/XML, there is JSON/XML Schema for this purpose. In case of JSON-LD, JSON Schema is not sufficient. Therefore, so-called application profiles are used, which are defined by using Shapes Constraint Language (SHACL).
In general, the Turntable API is split in two parts:
The first part is about managing (CRUD) 'schemas/application profiles'which are describing the structure of metadata documents, extended by the ability to validate a metadata document against a registered 'schema/application profile/⋯'.
The second part is about managing the metadata documents itself based on one of registered 'schemas/application profiles'.
Currently, we build the Metadata Hub as a Demonstrator mapping two completely different repositories (Coscine, MetaStore) as a showcase. The Metadata Hub is powered by the Metadata Hub Framework and provides a web interface to make the service available to a broad mass.
This work has been supported by the research program 'Engineering Digital Futures'of the Helmholtz Association of German Research Centers, the Helmholtz Metadata Collaboration Platform and the German National Research Data Infrastructure (NFDI).

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify ”other” (stakeholder)

In addition please add keywords.

FAIR, services, tools, metadata

Primary author: HARTMANN, Volker (KIT)

Co-author: Mr HEINRICHS, Benedikt (RWTH Aachen)

Presenter: HARTMANN, Volker (KIT)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **68** Contribution code: **1-15** Type: **Poster**

Intergrated Data Workflow using HELIPORT at TELBE

At the High-Field High-Repetition-Rate Terahertz facility @ ELBE (TELBE)1,
ultrafast terahertz-induced dynamics can be probed in various states of matter with highest precision. The TELBE sources offer both, stable and tunable narrowband THz radiation with pulse energies of several microjoules at high repetition rates and a synchronized coherent diffraction radiator,that provides broadband single-cycle pulses. The measurements at TELBE are data intensive, which can be as high as 20GB per experiment, that can lasts up to several minutes. As a result, the current data aquisition and data analysis stages are decoupled, where in the first step the primary data is processed and stored at HZDR and in a later step, restricted data access is made available to the user for post-processing.

In this poster contribution, we present an integrated workflow for post-processing of the experimental data at TELBE with in-built exchange of metadata between the experiment control software LabView and the workflow execution engine UNICORE2. We also present the guidance system HELIPORT3 which manages the metadata of the associated project proposal and job information from UNICORE, and integrates with the electronic lab notebook (MediaWiki4), providing a user-friendly interface for monitoring the actively running experiments at TELBE.

1 https://doi.org/10.1063/1.4978042
2 https://doi.org/10.1109/HPCSim.2016.7568392
3 https://doi.org/10.1145/3456287.3465477
4 https://www.mediawiki.org/wiki/Project:About

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Producer

Please specify ”other” (stakeholder)

In addition please add keywords.

Intergrated Data Workflow
HELIPORT
UNICORE

Primary authors: PAPE, David (HZDR); LOKAMANI, Mani (HZDR); JUCKELAND, Guido (Helmholtz-Zentrum Dresden-Rossendorf); DEINERT, Jan-Christoph (Helmholtz-Zentrum Dresden-Rossendorf); KELLING,

Jeffrey (HZDR); VOIGT, Martin (HZDR); KNODEL, Oliver (Helmholtz-Zentrum Dresden-Rossendorf); MUELLER, Stefan (Helmholtz-Zentrum Dresden-Rossendorf); GRUBER, Thomas (HZDR)

Presenter: LOKAMANI, Mani (HZDR)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **69** Contribution code: **1-13** Type: **Poster**

Toward a digital twin at the NeXus file level

Here, we report on our approach to establish a durable, rigid connection between the Aquarius beamline at synchrotron source Bessy II and its digital counterpart build in the simulation software Ray-UI 1. While simulations play a crucial role in the instrument design as a digital precursor of the real-world object and contain a comprehensive description of the setup, usually the digital representation is neglected once the real instrument is fully commissioned.
To preserve the symbiosis of simulated and real-world instrument beyond commissioning and approach the digital twin concept we combine the two worlds at the NeXus file level 2. For this purpose, the instrument section of the NeXus file is enriched by detailed simulation parameters where the current state of the instrument is reflected by including real motor positions, e. g. to incorporate the actual aperture of a slit system. As a result, on one hand, the enriched instrument description increases the reusability of experimental data in sense of the FAIR principles 3 and, on the other hand, allows to perform simulations of a measurement from the NeXus file, ready to be exploited by machine-learning techniques, e. g. for predictive maintenance.

1 P. Baumgärtel, P. Grundmann, T. Zeschke, A. Erko, J. Viefhaus, F. Schäfers, and H. Schirmacher, RAY-UI: New Features and Extensions, AIP Con. Proc. 2054, 060034 (2019).
2 https://manual.nexusformat.org/index.html
3 Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Please assign your poster to one of the following keywords.

Tools

Please assign yourself (presenting author) to one of the stakeholders.

Data Curators

Please specify ”other” (stakeholder)

In addition please add keywords.

Digital Twin, Simulation, NeXus, Instrumentation

Primary authors: GÜNTHER, Gerrit (Helmholtz-Zentrum Berlin); MANNIX, Oonagh (Helmholtz-Zentrum Berlin für Materialien und Energie GmbH (HZB)); BAUMGÄRTEL, Peter (Helmholtz-Zentrum Berlin für Materialien und Energie GmbH (HZB)); VADILONGA, Simone (Helmholtz-Zentrum Berlin für Materialien und Energie GmbH (HZB))

Presenter: GÜNTHER, Gerrit (Helmholtz-Zentrum Berlin)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **70** Contribution code: **1-37** Type: **Poster**

Pilot Dashboard for Open and FAIR Data Metrics by HMC Hub Matter

Making research data reusable in an open and FAIR 1 way is part of good scientific practice and is increasingly becoming part of the scientific workflow. Where and how "FAIR" research data is published alongside a research paper, is often not tracked by research institutes. In a pilot project of the Helmholtz Metadata Collaboration (HMC) Hub Matter we developed an approach to automatically find and catalogue publicly accessible datasets published by researchers of selected Helmholtz centers. These datasets are assessed with respect to the FAIR data guidelines using the F-UJI tool. [2,3] The results are gathered and visualized in an interactive pilot dashboard. This assists HMC Hub Matter to identify and characterize repositories used by the Matter community and to identify key actions for engaging with repository infrastructure and research communities. In this poster, we discuss the different steps of the data collection and the first results.
1 https://doi.org/10.1038/sdata.2016.18
2 https://doi.org/10.5281/zenodo.4063720
3 https://doi.org/10.5281/zenodo.6461229

Please assign your poster to one of the following keywords.

other

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify "other" (stakeholder)

HMC Hub Matter

In addition please add keywords.

Data-Management, Data-Mining, FAIR, Matter

Primary authors: GILEIN, Astrid (Helmholtz-Zentrum Berlin für Materialien und Energie); WALTER, Konstantin Pascal (Helmholtz-Zentrum Berlin für Materialien und Energie)

Co-authors: GLODOWSKI, Tempest (Helmholtz-Zentrum Berlin für Materialien und Energie); PREUSS, Gabriel (Helmholtz-Zentrum Berlin für Materialien und Energie); SCHMIDT, Alexander (Helmholtz-Zentrum Berlin für Materialien und Energie); SERVE, Vivien (Helmholtz-Zentrum Berlin für Materialien und Energie); MANNIX, Oonagh (Helmholtz-Zentrum Berlin für Materialien und Energie GmbH (HZB)); KUBIN, Markus (HMC, HZB)

Presenters: GILEIN, Astrid (Helmholtz-Zentrum Berlin für Materialien und Energie); WALTER, Konstantin Pascal (Helmholtz-Zentrum Berlin für Materialien und Energie)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **71** Contribution code: **2-13** Type: **Poster**

Meta-analysis of positive controls and laboratory metainformation in microbiome data

Recent advances in next-generation deep sequencing technologies have revolutionized our understanding of the microbiota's contribution to human health and disease. However, there are as many microbiome-disease associations as there are different protocols for generating microbiome data. This heterogeneity in laboratory data generation methods leads to protocol-specific biases in microbiome data and limits the comparability of microbiome studies. The biases can potentially be quantified by evaluating positive controls, i.e. microbiome mock communities with known sample composition that are processed along with biological samples.
We aim to build a database of published microbiome studies that used standardized, commercially available positive controls, and collect the studies'laboratory metadata to quantify the impact of different laboratory methods on microbiome data.
Therefore, we performed a systematic literature search of scientific papers using commercially available positive controls, and performed an initial meta-analysis for one mock community.
The pilot mock for meta-analysis, MSA-2002, was mentioned in 32 articles, of which seven remained for collection of lab metadata after applying exclusion criteria. On average, these seven studies provided 12 (median) out of 22 required laboratory metainformation factors. Combining the mock sequencing data of a subset of five studies revealed the substantial impact of study-specific biases on microbiome results.
Further analysis of the remaining mock communities is needed to assess whether positive controls can be used to quantify the biases introduced by laboratory methods. Our pilot project has shown that many scientific articles do not provide the necessary laboratory information to understand and reproduce their data. Moreover, the relevant pieces of information are often inconsistently described or scattered across the methods section, requiring automated paper scraping methods to extract them. The field of microbiome research needs to advance its reporting standards for laboratory metainformation to ensure reproducibility and comparability of microbiome data.

Please assign your poster to one of the following keywords.

Standards

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Re-User

Please specify ”other” (stakeholder)

In addition please add keywords.

Meta-analysis, laboratory metadata, reporting standards

Primary authors: RAUER, Luise; GENTZ, Tanja; KARBHAL, Rajiv; HÜLPÜSCH, Claudia; REIGER, Matthias; TRAIDL-HOFFMANN, Claudia; MÜLLER, Christian L.; NEUMANN, Avidan U.

Presenter: RAUER, Luise

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **72** Contribution code: **2-18** Type: **Poster**

Metadata curation use cases in astroparticle physics

Demanding requirements of fundamental physics at large-scale facilities are forcing researchers to use and further develop sophisticated computer science for high-efficient data processing, analysis, curation and preservation. PUNCH4NFDI (Particles, Universe, NuClei and Hadrons for the NFDI) is a consortium of particle, astroparticle, astro-, hadron, and nuclear physics, looking forward to developing advanced techniques and concepts for scientific big data. An important part of these developments represents in-depth studies of best practices of big data access and transfer, as well as adaptation of effective metadata curation strategies.
Prerequisites for development of a user-level metadata schema include a deep knowledge of all the peculiarities of the heterogeneous data supplied to the system from various distributed data sources, as well as a comprehension of the relevant user experiences and the necessary system functionality. Moreover, there is a significant variety in the practices of working with data and research conduction within the consortium. In this regard, study of user scenarios within individual research groups is of particular importance.
In this contribution, a comparative analysis of two metadata curation use cases from the PUNCH4NFDI consortium will be presented. We will consider the experience of two projects in the field of astroparticle physics, KASCADE Cosmic-ray Data Center (KCDC) and German-Russian Astroparticle Data Life Cycle Initiative (GRADLCI) in the context of the aims and requested functionality, chosen data architectures, technical solutions and, especially, metadata management approaches.
Acknowledgement: This work was partially supported by DFG fund „NFDI 39/1“for the PUNCH4NFDI consortium.

Please assign your poster to one of the following keywords.

other

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify ”other” (stakeholder)

In addition please add keywords.

astroparticle physics, metadata, open science

Primary author: TOKAREVA, Victoria (KIT)

Presenter: TOKAREVA, Victoria (KIT)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **73** Contribution code: **2-01** Type: **Poster**

Development of a metadata schema for publication of health-related research data on the German Central Health Study Hub of the NFDI4Health

Introduction: The National Research Data Infrastructure for Personal Health Data (NFDI4Health) aims to improve the FAIRness of health-related data from epidemiological, public health and clinical studies as well as registries and administrative health databases1. One key service of NFDI4Health is the German Central Health Study Hub2 that supports a standardised publication and search of research (meta-)data (i.e. study-level data and related documents3) and is based on a metadata schema developed by NFDI4Health[4, 5].
Concept: The NFDI4Health metadata schema allows to collect basic bibliographic information such as title and description as well as information about related persons, organisations and publications. Additionally, details about study design and data accessibility can be provided. The DataCite Metadata Schema6 was taken as a basis for the schema as it utilises a broadly applied vocabulary and supports publication of research data including DOI assignment. To describe health studies, the schema was extended by attributes from well-established data models in clinical and population-based research[7, 8, 9, 10, 11]. Additionally, a mapping of the schema against the HL7®FHIR® standard was conducted to evaluate its compatibility and to prepare profiling[12].
Conclusion and outlook: The NFDI4Health metadata schema provides a structured way to bundle metadata of various research data types from different health-related domains. The incorporation of a generic metadata standard enables description, publication and comparison of health studies and related resources at a general level. At the same time, the schema is scalable to adjacent research fields such as social sciences. Furthermore, domain-specific attributes captured in the schema for health research allow access to more in-depth information. Currently, the schema is being expanded to meet the needs of specific use cases, i.e. nutritional epidemiology, epidemiology of chronic diseases, and secondary data and record linkage. Mappings to other health metadata schemas such as ECRIN[13] and ERDRI[14] are also intended.

Please assign your poster to one of the following keywords.

Standards

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify ”other” (stakeholder)

In addition please add keywords.

Metadata model; FAIR principles; Health

Primary author: SHUTSKO, Aliaksandra (ZB MED –Information Centre for Life Sciences)

Co-authors: SCHMIDT, Carsten Oliver (Institute of Community Medicine, University Greifswald); KLOPFENSTEIN, Sophie Anne Ines (Berlin Institute of Health (BIH) at Charité –Universitätsmedizin Berlin); LÖBE, Matthias (Institute for Medical Informatics (IMISE), University of Leipzig); ABAZA, Haitham (Heidelberg Institute for Theoretical Studies (HITS)); GOLEBIEWSKI, Martin (Heidelberg Institute for Theoretical Studies (HITS)); DARMS, Johannes (ZB MED –Information Centre for Life Sciences); FLUCK, Juliane (ZB MED –Information Centre for Life Sciences)

Presenter: SHUTSKO, Aliaksandra (ZB MED –Information Centre for Life Sciences)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **74** Contribution code: **2-32** Type: **Poster**

A Digital Research Process for FAIR Data and Metadata

With new specialisations such as Data Science driven by digitisation, efficiency potentials of a digital transformation are raised in both empirical research and data governance processes. Here, one challenge is to establish open and interoperable datasets, recognising the FAIR criteria (cf. Wilkinson et al., 2016) as a standard of that process. Data –as well as metadata –should comply to this standard. However, traditional methodological research processes (cf. Brosius, Haas, & Koschel, 2012, p. 28; Friedrichs, 1990, p. 119) lack the support of information technology which would lever the process into the digital age. Therefore, we propose a digital research process that closes ranks between the traditional process and the opportunities of a digital world.
The digital research process was established as a concept for a data model with corresponding roles (cf. Wuchner, & Sautter, 2020; Sautter, & Wuchner, 2020; Sautter et al., 2018). We found that a data governance process depends less on the specific method and much more on a common cross-method research process (cf. also UK Data Service). As a result, the digital research process needed to be highly adaptive to the purposes of different kinds of research fields.
The digital research process we propose consists of nine activities, terminated by data filing points (DFPs). The obligatory DFPs consider projects that focus on data search, acquisition, and archiving only. The optional DFPs represent the process of obtaining new (research) data. Additionally, optional data analysis may play a role in projects that merely reuse existing data. The optional DFPs represent the adaptability of research objectives in humanities. Equally unique to the digital research process is the frequent update of metadata throughout the research cycle, to create FAIR metadata throughout the time frame of the research and data processing.

Please assign your poster to one of the following keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Please specify ”other” (stakeholder)

In addition please add keywords.

Primary author: ANNIÉS, Jeannette (Institute of Human Factors and Technology Management IAT, University of Stuttgart)

Co-authors: SAUTTNER, Johannes (Institute for Industrial Engineering IAO, Fraunhofer-Gesellschaft); WUCHNER, Andrea (Information Center for Planning and Building IRB, Fraunhofer-Gesellschaft); DOBROKHOTOVA, Ekaterina (Institute for Industrial Engineering IAO, Fraunhofer-Gesellschaft)

Presenter: ANNIÉS, Jeannette (Institute of Human Factors and Technology Management IAT, University of Stuttgart)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **75** Contribution code: **2-26** Type: **Poster**

Helmholtz Digitization Ontology (HDO): harmonized descriptions of digital assets and processes to support the integrity of the Helmholtz digital ecosystem

The Helmholtz digital ecosystem connects diverse scientific domains with differing (domain-specific) standards and best practices for handling metadata. Ensuring interoperability within such a system, e.g. of developed tools, offered services and circulated research data, requires a semantically harmonized, machine-actionable, and coherent understanding of the relevant concepts. Further, this needs to be aligned and harmonized with European and global initiatives to ensure an open and interoperable flow of data and information.
Accordingly, the Helmholtz Metadata Collaboration develops the "Helmholtz Digitization Ontology"(HDO), which contains machine-actionable descriptions of digital assets and processes relevant to this ecosystem. Containing consistent and carefully curated semantics, it is intended to serve as an institutional reference thereby supporting the integrity of HMC developments internally as well as externally.
HDO is aligned to practices and conventions of the Open Biological and Biomedical Ontologies (OBO)1: we produce definitions in the OBO recommended genus-differentia form (i.e. for each term we define a Genus as well as its Differentia) that are coherent and precise. Class labels and definitions are developed bilingually in English and German and further contain information on synonymy, comments as well as micro-credits of contributions. The HDO is implemented based on the Ontology Development Kit (ODK)2 to ensure long-term development. The current development status can be followed in our public git repository3 –a 1st release of the HDO is planned in Q3/Q4 2022.
1 https://obofoundry.org/
2 https://arxiv.org/abs/2207.02056
3 https://gitlab.hzdr.de/hmc/hmc-public/hob/hdo

Please assign your poster to one of the following keywords.

Semantics

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify "other" (stakeholder)

In addition please add keywords.

Semantics, Ontology, Helmholtz, Harmonization, Interoperability

Primary authors: BUTTIGIEG, Pier Luigi (GEOMAR Helmholtz-Zentrum für Ozeanforschung, Kiel, Germany); FATHALLA, Said; GUENTHER, Gerrit (Helmholtz-Zentrum Berlin); HOFMANN, Volker; LEHMANN, Jos (German Cancer Research Center (Deutsches Krebsforschungszentrum - DKFZ), Heidelberg, Germany); STEINMEIER, Leon (Helmholtz Institute Freiberg); VIDEGAIN BARRANCO, Pedro (Forschungszentrum Jülich)

Co-author: LEMSTER, Christine (Geomar)

Presenters: BUTTIGIEG, Pier Luigi (GEOMAR Helmholtz-Zentrum für Ozeanforschung, Kiel, Germany); FATHALLA, Said; GUENTHER, Gerrit (Helmholtz-Zentrum Berlin); HOFMANN, Volker; LEHMANN, Jos (German Cancer Research Center (Deutsches Krebsforschungszentrum - DKFZ), Heidelberg, Germany); STEINMEIER, Leon (Helmholtz Institute Freiberg); VIDEGAIN BARRANCO, Pedro (Forschungszentrum Jülich); LEMSTER, Christine (Geomar)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **77** Contribution code: **2-22** Type: **Poster**

AquaDiva MetaData

The Collaborative Research Centre AquaDiva is a large collaborative project spanning a variety of domains, such as biology, geology, chemistry and computer science with the common goal to better understand the Earth's critical zone, in particular, how environmental conditions and surface properties shape the structure, properties, and functions of the subsurface. Within AquaDiva large volumes of heterogeneous observational data are being collected. Besides this structured data, knowledge is also encoded in an unstructured form in scientific publications. To support search and dataset discovery, standard metadata is recommended to describe data. However, and due to the heterogeneity in AquaDiva datasets, one metadata standard does not fit all. In the first phase, we made use of EML and ABCD, however, both of them are not adequate for AquaDiva. Therefore, we develop and introduce AquaDiva specific metadata to effectively describe AquaDiva data and support dataset search and discovery. In particular, the proposed metadata consists of four main components: a general component to provide information about the dataset, such as title, description and a set of main keywords; the second component to introduce information about the project(s) involved in collecting and generating the dataset; the third component to describe information about persons, such as dataset owner, dataset curators; the last and the most important component to present AquaDiva-specific metadata information, such as sampling location, sample object, sample type and the data types generated from these samples. As a next step, we plan to link our metadata concepts (or to annotate our metadata concepts with) to appropriate controlled vocabularies and ontologies. This does not only contribute to interoperability, but also ensures a well understood, common definition of the fields.

Please assign your poster to one of the following keywords.

Semantics

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify "other" (stakeholder)

Scientist / Data Management

In addition please add keywords.

Primary author: ALGERGAWY, Alsayed (Heinz-Nixdorf Chair for Distributed Information Systems, Friedrich Schiller University, Jena, Germany)

Co-authors: HAMED, Hamdi (Heinz-Nixdorf Chair for Distributed Information Systems, Friedrich Schiller University, Jena, Germany); THIEL, Sven (Alsayed); KÖNIG-RIES, Birgitta (Heinz-Nixdorf

Chair for Distributed Information Systems, Friedrich Schiller University, Jena, Germany)

Presenter: ALGERGAWY, Alsayed (Heinz-Nixdorf Chair for Distributed Information Systems, Friedrich Schiller University, Jena, Germany)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **78** Contribution code: **2-15** Type: **Poster**

Medical Imaging as a Case Study of the Use of Metadata in Health Research Data Management

The Helmholtz Metadata Collaboration (HMC) promotes the use of metadata in Research Data Management as a means to achieving data findability, accessibility, interoperability, reusability (FAIR). These in turn enable or optimize software functionalities essential to automated research processes, such as multi-, inter- and transdisciplinary indexing and retrieval, versioning, provenance tracking, data contextualization, workflow reproduction, compliance assessment, publication. Metadata are also key to Hybrid Artificial Intelligence, i.e. the integration of sub-symbolic and symbolic techniques, which improves Machine Learning systems'trainability and explainability.

In this context, Hub Health is developing a framework for the identification and specification of metadata use cases in health data analysis workflows. This will provide stakeholders, e.g. researchers or developers, with insight into types and roles of metadata in a given workflow phase. For instance, metadata that are more extrinsic to the data, such as data format, are mostly needed during data acquisition. Metadata that are more intrinsic to the data, such as terminologies, may also contribute to the data analysis itself, for instance during feature extraction.

The initial case study for the framework is Medical Imaging. This offers a template for workflows that are ubiquitous in medical research and practice (e.g. diagnostics and prognostics) and it can be extended to the analysis of data other than images, e.g. natural language or diagnostic test data.

The toolkit KAAPANA, which supports AI-based medical data analysis workflows, is being used to benchmark the use of metadata in medical imaging workflows, to test new ideas and to integrate HMC resources, such as the Hub Health Information Portal or other tools from FAIR Data Commons or other Hubs.
In particular, image segmentation in KAAPANA, a phase shared by many imaging workflows, is currently being reviewed from the perspective of the functionalities targeted in HMC (from indexing and retrieval to trainability and explainability).

Please assign your poster to one of the following keywords.

other

In addition please add keywords.

Standards, Semantics, Tools

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify ”other” (stakeholder)

Metadata Scientist

Primary authors: LEHMANN, Jos (German Cancer Research Center (Deutsches Krebsforschungszentrum - DKFZ), Heidelberg, Germany); SCHADER, Philipp (DKFZ); KULLA, Lucas (DKFZ); NOLDEN, Marco (DKFZ); MAIER-HEIN, Klaus (DKFZ)

Presenter: LEHMANN, Jos (German Cancer Research Center (Deutsches Krebsforschungszentrum - DKFZ), Heidelberg, Germany)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **79** Contribution code: **2-04** Type: **Poster**

HARMONise –Enhancing the interoperability of marine biomolecular (meta)data across Helmholtz Centres

Biomolecules, such as DNA and RNA, provide a wealth of information about the distribution and function of marine organisms, and biomolecular research in the marine realm is pursued across several Helmholtz Centers. Biomolecular metadata, i.e. DNA and RNA sequences and all steps involved in their creation, exhibit great internal diversity and complexity. However, high-quality (meta)data management is not yet well developed and harmonized in environmentally focused Helmholtz Centers. As part of the HMC Project HARMONise, we aim to develop sustainable solutions and digital cultures to enable high-quality, standards-compliant curation and management of marine biomolecular metadata at AWI and GEOMAR to better embed biomolecular science into broader digital ecosystems and research domains. Our approach builds on a relational database that aligns metadata with community standards such as the MIxS (Minimum Information about any (x) sequence) supported by the International Nucleotide Sequence Database Collaboration (INSDC) to promote global interoperability. At the same time, we ensure the harmonization of metadata with existing Helmholtz repositories (e.g. PANGAEA). A web-based hub will enable the standardized export and exchange of core metadata, in line with domain-specific standards and using standard conventions such as JSON(-LD). Here we will present the current status of the database scheme, highlight the use of standards and fields that promote interoperability, and outline the planned development of an exchange hub for sharing and validating biomolecular metadata across Helmholtz Centers. Enabling sustainable data stewardship, export and publication routines will support researchers in delivering Helmholtz biomolecular data to national European and global repositories in alignment with community standards and the FAIR principles.

Please assign your poster to one of the following keywords.

Standards

In addition please add keywords.

sequence-data-management, interoperability, metadata-harmonization, FAIR-principles

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Producer

Please specify ”other” (stakeholder)

Primary author: BIENHOLD, Christina (AWI Helmholtz Centre for Polar and Marine Research)

Co-authors: HARMS, Lars; KOPPE, Roland; NEUHAUS, Stefan; BAYER, Till

Presenter: BIENHOLD, Christina (AWI Helmholtz Centre for Polar and Marine Research)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **80** Contribution code: **1-22** Type: **Poster**

ALAMEDA –A scalable multi-domain metadata management platform

Modern Earth sciences produce a continuous increasing amount of data. These data consist of the measurements/observations and descriptive information (metadata) and include semantic classifications (semantics). Depending on the geoscientific parameter, metadata are stored in a variety of different databases, standards and semantics, which is obstructive for interoperability in terms of limited data access and exchange, searchability and comparability. Examples of common data types with very different structure and metadata needs are maps, geochemical data derived from field samples, or time series data measured with a sensor at a point, such as precipitation or soil moisture. So far, there is a large gap between the capabilities of databases to capture metadata and their practical use. ALAMEDA is designed as modular structured metadata management platform for curation, compilation, administration, visualization, storage and sharing of meta information of lab-, field- and modelling datasets. As a pilot application for stable isotope and soil moisture data ALAMEDA will enable to search, access and compare meta information across organization-, system- and domain boundaries. ALAMEDA covers 5 major categories: observation & measurements, sample & data history, sensor & devices, methods & processing, environmental characteristics (spatio & temporal). These categories are hierarchically structured, interlinkable and filled with specific metadata attributes (e.g. name, data, location, methods for sample preparation, measuring and data processing, etc.). For the pilot, all meta information will be provided by existing and wellestablished data management tools (e.g. mDIS, Medusa, etc.). In ALAMEDA, all information is brought together and will be available via web interfaces. Furthermore, the project focuses on features such as metadata curation with intuitive graphical user interfaces, the adoption of well-established standards, the use of domain-controlled vocabularies and the provision of interfaces for a standards-based dissemination of aggregated information. Finally, ALAMEDA should be integrated into the DataHub (Hub-Terra).

Please assign your poster to one of the following keywords.

Tools

In addition please add keywords.

Metadata, reusability, availability

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Producer

Please specify ”other” (stakeholder)

Presenter: RACH, Oliver

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **81** Contribution code: **2-07** Type: **Poster**

ADVANCE: Advanced metadata standards for biodiversity survey and monitoring data for supporting research and conservation

In an ever-changing world, field surveys, inventories and monitoring data are essential for prediction of biodiversity responses to global drivers such as land use and climate change. This knowledge provides the basis for appropriate management. However, field biodiversity data collected across terrestrial, freshwater and marine realms are highly complex and heterogeneous. The successful integration and re-use of such data depends on how FAIR (Findable, Accessible, Interoperable, Reusable) they are. ADVANCE aims at underpinning rich metadata generation with interoperable metadata standards using semantic artefacts. These are tools allowing humans and machines to locate, access and understand (meta) data, and thus facilitating integration and reuse of biodiversity monitoring data across terrestrial, freshwater and marine realms. To this end, we revised, adapted and expanded existing metadata standards, thesauri and vocabularies. We focused on the most comprehensive database of biodiversity monitoring schemes in Europe (DaEuMon) as the base for building a metadata schema that implements quality control and complies with the FAIR principles. In a further step, we will use biodiversity data to test, refine and illustrate the strength of the concept in cases of real use. ADVANCE thus complements semantic artefacts of the Hub Earth & Environment and other initiatives for FAIR biodiversity research, enabling assessments of the relationships between biodiversity across realms and associated environmental conditions. Moreover, it will facilitate future collaborations, joint projects and data-driven studies among biodiversity scientists of the Helmholtz Association and beyond.

Please assign your poster to one of the following keywords.

Standards

In addition please add keywords.

biodiversity monitoring, metadata standards, FAIR

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Re-User

Please specify ”other” (stakeholder)

Primary author: SILVA MENGER, Juliana (UFZ, AWI)

Presenter: SILVA MENGER, Juliana (UFZ, AWI)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **82** Contribution code: **2-12** Type: **Poster**

enhancing FAIRness in seismological data management

Within the current project we plan to optimise data and metadata curation workflow by automating the creation of community standard metadata StationXML and include the generated PIDs as well as link them to the parent dataset DOIs. Moreover, we plan to enrich metadata with terms from standard and community specific vocabularies. Specific guidelines, describing the OBS data management workflows, have been developed during the first part of the project; data from AWI and GEOMAR are being archived at the GFZ EIDA node using these guidelines. Metadata have been created starting from the library of instruments at AWI and using the OBSinfo tools developed at IPGP. A FAIR assessment is also being carried out for the datasets archived and exposed to the community. Leveraging on the experience gained at GFZ in the recent years, and taking advantage of the link to OBS communities at AWI and GEOMAR, we envisage a fully FAIR data management process acting as a blueprint for the Earth and Environment research community in general, within and beyond the Helmholtz Association. The workflow applied here is developed in synergy with the international seismological community within FDSN and ORFEUS.

Please assign your poster to one of the following keywords.

Standards

In addition please add keywords.

seismology, OBS, PID, vocabulary, ontology

Please assign yourself (presenting author) to one of the stakeholders.

Please specify ”other” (stakeholder)

Data Infrastructure Provider and Data Curators

Primary authors: HILLMANN, Laura (laura@gfz-potsdam.de); HEMMLEB, S.; STROLLO, Angelo (GFZ); QUINTEROS, J.; HEINLOO, A.; HABERLAND, C.; HAXTER, M.; SCHMIDT-AURSCH, M.; ULMER, L.; DANNOWSKI, A.; KOPP, H.

Presenters: HILLMANN, Laura (laura@gfz-potsdam.de); STROLLO, Angelo (GFZ)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **83** Contribution code: **1-08** Type: **Poster**

HELIPORT —An Integrated Research Data Lifecycle

HELIPORT is a data management solution that aims at making the components and steps of the entire research experiment's life cycle discoverable, accessible, interoperable and reusable according to the FAIR principles.
Among other information, HELIPORT integrates documentation, scientific workflows, and the final publication of the research results - all via already established solutions for proposal management, electronic lab notebooks, software development and devops tools, and other additional data sources. The integration is accomplished by presenting the researchers with a high-level overview to keep all aspects of the experiment in mind, and automatically exchanging relevant metadata between the experiment's life cycle steps.
Computational agents can interact with HELIPORT via a REST API that allows access to all components, and landing pages that allow for export of digital objects in various standardized formats and schemas. An overall digital object graph combining the metadata harvested from all sources provides scientists with a visual representation of interactions and relations between their digital objects, as well as their existence in the first place. Through the integrated computational workflow systems, HELIPORT can automate calculations using the collected metadata.
By visualizing all aspects of large-scale research experiments, HELIPORT enables deeper insights into a comprehensible data provenance with the chance of raising awareness for data management.

Please assign your poster to one of the following keywords.

Tools

In addition please add keywords.

Data management, FAIR, workflows

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify ”other” (stakeholder)

Primary authors: PAPE, David (HZDR); KNODEL, Oliver (Helmholtz-Zentrum Dresden-Rossendorf)

Presenters: PAPE, David (HZDR); KNODEL, Oliver (Helmholtz-Zentrum Dresden-Rossendorf)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **84** Contribution code: **2-16** Type: **Poster**

Metadata generation, enrichment and linkage across the three domains health, environment and earth observation

Cross-domain research is often hampered by the lack of harmonized metadata schemas and standards. Metadata of different domains vary in origin, format and scope, so they cannot be merged routinely. In the interdisciplinary field of environmental epidemiology, an efficient linkage of health data with the multitude of environmental and earth observation data is crucial to quantify human exposures. To bridge the gap between the metadata of our different research fields, we aim to compile machine-readable and interoperable metadata schemas for exemplary data of our three domains Health (HMGU), Earth & Environment (UFZ), and Aeronautics, Space & Transport (DLR).

As use cases for metadata compilation, enrichment, and pooling, the project partners contributed with metadata of the child cohorts GINI and LISA (HMGU), drought monitor (UFZ) and land cover (DLR). UFZ and DLR will adopt as common standard ISO 19115: Geographic Metadata Information. For HMGU, we reviewed several metadata standards for health data (e.g. CDISC ODM, Snomed CT, HL7 FHIR) and started to upload our metadata to the NFDI4health StudyHub, an inventory of German health studies on COVID-19. In addition, we have developed a workflow to transform base cohort information in an ISO 19115 compliant manner without exposing sensitive information about participants'data.

Spatial and time coverage will be the main mapping criteria. The metadata of our three domains will be uploaded into an instance of GeoNetwork, a catalog application that we are currently setting up in a testing environment, where they can be jointly queried and searched. We aim to have a server version ready by the end of the project that can be augmented with additional metadata from our domains, but also from other fields, to facilitate interdisciplinary research.

Please assign your poster to one of the following keywords.

other

In addition please add keywords.

Health, environment, earth observation, mapping

Please assign yourself (presenting author) to one of the stakeholders.

Please specify ”other” (stakeholder)

Data Infrastructure Provider and Data Curator

Primary author: DALLAVALLE, Marco (HMGU /LMU)

Presenter: DALLAVALLE, Marco (HMGU /LMU)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **85** Contribution code: **1-28** Type: **Poster**

Automated FAIR4RS software publication with HERMES

Software as an important method and output of research should follow the RDA "FAIR for Research Software Principles". In practice, this means that research software, whether open, inner or closed source, should be published with rich metadata to enable FAIR4RS. For research software practitioners, this currently often means following an arduous and mostly manual process of software publication. HERMES, a project funded by the Helmholtz Metadata Collaboration, aims to alleviate this situation. We develop configurable, executable workflows for the publication of rich metadata for research software, alongside the software itself. These workflows follow a push-based approach: they use existing continuous integration solutions, integrated in common code platforms such as GitHub or GitLab, to harvest, unify and collate software metadata from source code repositories and code platform APIs. They also manage curation of unified metadata, and deposits on publication platforms. These deposits are based on deposition requirements and curation steps defined by a targeted publication platform, the depositing institution, or a software management plan. In addition, the HERMES project works to make the widely-used publication platforms InvenioRDM and Dataverse "research software-ready", i.e., able to ingest software publications with rich metadata, and represent software publications and metadata in a way that supports findability, assessability and accessibility of the published software versions. Beyond the open source workflow software, HERMES will openly provide templates for different continuous integration solutions, extensive documentation, and training material. Thus, researchers are enabled to adapt automated software publication quickly and easily. Our poster presents a high-level overview of the HERMES concept, its status and an outlook.

Please assign your poster to one of the following keywords.

Tools

In addition please add keywords.

Software publication, Software metadata, automation

Please assign yourself (presenting author) to one of the stakeholders.

Please specify "other" (stakeholder)

Scientist/Software Producer, Data/Software Infrastructure Provider, Metadata Curators

Primary author: DRUSKAT, Stephan (German Aerospace Center (DLR))

Presenter: DRUSKAT, Stephan (German Aerospace Center (DLR))

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **86** Contribution code: **1-16** Type: **Poster**

The HMC project Metamorphoses - Metadata for the merging of diverse atmospheric data on common subspaces

This poster presents the new HMC project Metamorphoses ("Metadata for the merging of diverse atmospheric data on common subspaces"). The project will develop enhanced standards for storage efficient decomposed arrays and tools for an automated generation of standardised Lagrange trajectory data files thus enabling an optimised and efficient synergetic merging of large remote sensing data sets. We detail the individual objectives of the project and show first results of merging data from two different satellite sensors.

Please assign your poster to one of the following keywords.

Tools

In addition please add keywords.

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Producer

Please specify "other" (stakeholder)

Primary author: SCHNEIDER, Matthias (Karlsruhe Institute of Technology)

Presenter: SCHNEIDER, Matthias (Karlsruhe Institute of Technology)

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **87** Contribution code: **2-11** Type: **Poster**

SECoP@HMC - Metadata in the Sample Environment Communication Protocol

The integration of sample environment (SE) equipment in a beam line experiment is a complex challenge both in the physical world and in the digital world. Different experiment control software offer different interfaces for the connection of SE equipment. Therefore, it is time-consuming to integrate new SE or to share SE equipment between facilities.

To tackle this problem, the International Society for Sample Environment (ISSE) developed the Sample Environment Communication Protocol (SECoP) to standardize the communication between instrument control software and SE equipment (see 1 and references therein). SECoP offers, on the one hand, a generalized way to control SE equipment. On the other hand, SECoP holds the possibility to transport SE metadata in a well-defined way.

Using SECoP as a common standard for controlling SE equipment and generating SE metadata will save resources and intrinsically give the opportunity to supply standardized and FAIR data compliant SE metadata. It will also supply a well-defined interface for user-provided SE equipment, for equipment shared by different research facilities and for industry.

In this presentation we will present the SECoP@HMC project supported by the Helmholtz Metadata Collaboration and give an overview of the present status.

Please specify ”other” (stakeholder)

Please assign your poster to one of the following keywords.

Standards

In addition please add keywords.

SECoP, Sample-Environment, Experiment-Control, Metadata-Standards

Please assign yourself (presenting author) to one of the stakeholders.

Scientist/ Data Producer

Primary author: KIEFER, Klaus

Co-authors: PETTERSON, A.; KLEMKE, B.; BRANDL, G.; ROSSA, L.; ZOLLIKER, M.; EKSTRÖM, N.; HERMANNSDÖRFER, T.; KRACHT, T.

Presenter: KIEFER, Klaus

Session Classification: Postersession II

Track Classification: Postersession

Contribution ID: **88** Contribution code: **2-28** Type: **Poster**

HMC Earth and Environment - Overall Strategy and Implementation of a FAIR Helmholtz Data Space

HMC Earth and Environment (E&E) strives to define, create and activate a Helmholtz FAIR Data Space (HFDS) as a "decentralized infrastructure for trustworthy data sharing and exchange in data ecosystems based on commonly agreed principles" (Nagel L., Lycklama D., 2021). Within HMC E&E the data space consists of common agreements to implement the FAIR building blocks (see below), leading to internal interoperability of data. In addition a data integration system is needed, which will act as a data broker between data infrastructures, providing internal and external integration and data access opportunities.

Unlike the concept for the European data space, which is largely tailored towards commercial data, the Helmholtz Association's data space covers primarily research data, which may or may not be openly accessible. For such research data, we envision four major building blocks required, to implement and activate the data space:

1. the consistent usage of high-resolution PID referable metadata elements, e.g. ORCID, DOI, ROR, InstPID and others (see other poster).
2. The implementation of consistent semantic concepts within data repositories and infrastructures (see other poster).
3. The containerization datasets and metadata within machine actionable FAIR digital objects (FDOs)
4. The agreement and implementation of standardized interfaces to access and address containerized data through common APIs.

To implement these features a co-design and implementation process needs to be set up. Within this co-design process procedures should be agreed upon, implemented and maintained by data stewards and data infrastructures together, in order to support data producers, data maintaners and data re-users and ease their handling of research data.

In HMC we plan to conduct the following actions, in order to develop the HFDS together with our scientific and technical communities:

1. Define the concept and requirements of a Helmholtz data space, which is in-line with other data spaces in preparation.
2. Establish a communication platform allowing us to define and agree upon the building blocks required to set up the data space (see above and other poster).
3. Work with data repositories and data stewards to implement and document the building blocks required to establish the data space.
4. Build a data integration system, connecting the different decentralized parts of the data space and connect it to other data spaces.

These activities will ultimately lead to the establishment of an well-formed interoperable FAIR Data Space, which anyone interested is welcome to join and shape.

(1) Nagel L., Lycklama D. (2021): Design Principles for Data Spaces. Position Paper. Version 1.0. Berlin, DOI: http://doi.org/10.5281/zenodo.5105744

Please assign your poster to one of the following keywords.

other

In addition please add keywords.

Strategy, Vision, Overview, Data Space

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify ”other” (stakeholder)

Data Infrastructure Provider, Data Curators

Primary authors: SÖDING, Emanuel (GEOMAR); PÖRSCH, Andrea (HMC Hub EE at GFZ); WEINELT, Martin (GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel); BUTTIGIEG, Pier Luigi (GEOMAR Helmholtz-Zentrum für Ozeanforschung, Kiel, Germany); LORENZ, Sören (GEOMAR Helmholtz Centre for Ocean Research Kiel)

Presenter: SÖDING, Emanuel (GEOMAR)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **89** Contribution code: **2-24** Type: **Poster**

HMC Earth and Environment - Semantics, connecting Helmholtz data with with international initiatives

In pursuit of deep and expressive semantic interoperability, the Earth and Environment Hub is adopting a three-pillared approach to develop strategically and technically aligned capacity within the Helmholtz Association and globally.

The first pillar is implementation of high-quality, future-oriented semantic solutions for Earth and environmental applications. HMC E&E personnel lead the development of the Environment Ontology (ENVO), an internationally recognised, highly expressive, and adopted community ontology for environmental research, management, and operations. Leveraging the practices, technologies, and interoperability architecture of the Open Biomedical and Biological Ontologies (OBO) Foundry, ENVO hosts machine-friendly representations of classification systems including the World Wildlife Fund's biomes and ecoregions, the Global Platform for Marine Litter's litter and debris classification for reporting towards Sustainable Development Goal (SDG) 14, and the UNEP World Conservation Monitoring Centre's mountain classification. Current activities are deepening links to the SDGs through collaboration with UN Environment, the UN Statistical Division, and UN Data, particularly on environmental hazards and disasters.

Our second pillar is harmonisation of existing semantic resources to enhance interoperability amongst them. Through the work of an Earth Science Information Partners' (ESIP) cluster for semantic harmonisation, we are supporting efforts to harmonise semantics for vocabularies, glossaries, thesauri, and ontologies describing the cryosphere, the marine realm, natural hazards and disasters, and heliosphere. This activity engages major global stakeholders - including the WMO and NASA - and fosters collaborative interoperation between formerly competing standards.

Our third pillar is the co-development and deployment of lightweight semantic solutions for knowledge graph creation and maintainence by multiple parties. Leveraging ESIP's Science on Schema (SoSo) approaches, our personnel are leading co-development of the UNESCO Intergovernmental Oceanographic Commission's Ocean Data and Information System (ODIS) and Ocean InfoHub (OIH), requested by the Member States. As it matures, we seek to merge this graph with its counterparts emerging in the Polar community and others, as well as ontological graphs noted in the other pillars.

In conclusion, our efforts are addressing local and global needs in environmental semantics through broad, multilateral collaboration while creating fluid capacity exchange between all actors. This approach will support the creation of Helmholtz data spaces bearing intrinsic semantic compatibility with external systems and ready to transfer Helmholtz data to address global challenges.

Please assign your poster to one of the following keywords.

Semantics

In addition please add keywords.

Strategy, Semantics, International Connections

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify ”other” (stakeholder)

Data Infrastructure Provider, Data Curators

Primary authors: BUTTIGIEG, Pier Luigi (GEOMAR Helmholtz-Zentrum für Ozeanforschung, Kiel, Germany); SÖDING, Emanuel (GEOMAR); PÖRSCH, Andrea (HMC Hub EE at GFZ); WEINELT, Martin (GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel); LORENZ, Sören (GEOMAR Helmholtz Centre for Ocean Research Kiel)

Presenters: BUTTIGIEG, Pier Luigi (GEOMAR Helmholtz-Zentrum für Ozeanforschung, Kiel, Germany); SÖDING, Emanuel (GEOMAR)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **90** Contribution code: **2-14** Type: **Poster**

HMC Earth and Environment - Using PIDs in Helmholtz Earth and Environment Data Infrastructures

PIDs (Persistent Identifiers) are a core concept at the center of FAIR data architectures such as FAIR Digital Objects. They point to a digital resource such as a publication, dataset or a set of information in a distinctive and lasting fashion and are assured to persist over longer, defined periods of time.

We looked into six established PID systems (ROR, ORCID, PIDINST, IGSN, DataCite DOI, Crossref DOI) to map the interconnection (graph) and overlap between systems. This was carried out by inspecting and comparing the metadata schemas of these PID systems in their current version to find out, to what extent they support each other as PID systems and how this is done.

We expected these PID schemas not to overlap in too many elements, but we expected some of the systems / schemas to recognize and make use of each another.

The number of external PID systems supported varies considerably for the six PID systems investigated, with ROR at 4 and up to 49 systems at ORCID. The system mostly implemented as a reference is DOI (4 other systems do refer to DOI in their metadata schema), while ROR is only referenced by DataCite DOIs yet.

Interconnected PID systems can be visualized as graphs of relationships (PID graphs) between for instance scientists, datasets, publications, institutions etc. They can be machine actionable and thus be tailored to specific questions or fields of interest, as was shown by EU programme FREYA (https://www.project-freya.eu).

Our findings show, that PIDs act as an important part of the data space we are constructing. They allow to link meta information of different data sets in a uniform manner. Consistently implementing PIDs will allow a high level of informational data interoperability among distributed data sets, which should complement other interoperability measures, e.g. the semantic interoperability.

Please assign your poster to one of the following keywords.

Standards

In addition please add keywords.

Strategy, Standards, PID, Earth, Environment

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify ”other” (stakeholder)

Data Infrastructure Provider, Data Curators

Primary authors: WEINELT, Martin (GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel); PÖRSCH, Andrea (HMC Hub EE at GFZ); BUTTIGIEG, Pier Luigi (GEOMAR Helmholtz-Zentrum für Ozeanforschung, Kiel, Germany); SÖDING, Emanuel (GEOMAR); LORENZ, Sören (GEOMAR Helmholtz Centre for Ocean Research Kiel)

Presenters: WEINELT, Martin (GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel); SÖDING, Emanuel (GEOMAR)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **91** Contribution code: **1-18** Type: **Poster**

HMC Earth and Environment - Community Involvement - the FAIR Implementation Network and the Community Portal

The desired Interoperability of data as outlined by the FAIR principles, requires a harmonization of data handling processes among data infrastructures. To support the adoption of agreements on such processes and thus further develop the "ROAD TO FAIR", HMC is currently establishing a FAIR-IMP-lementation Network (F-IMP). With this communication network we encourage the data management community to present suitable use cases, develop solutions and make recommendations for the implementation of data handling procedures not only within the Helmholtz Association but also beyond.

To activate the F-IMP a web-based information hub for all interested parties in the form of a communication portasl was set up. It aims to
enable a dialogue for the exchange of information leading to agreements across all participating parties, from Helmholtz Centers, over national initiatives to international actors, if desired. The platform will allow anyone to actively participate in discussions and decisions, not only the F-IMP, and track the ongoing coordination processes as they happen. The HMC Community Portal will initially be developed jointly by HMC and Helmholtz-Zentrum Hereon.

The following activities are currently planned to be conducted on the portal: Working Groups (WG) on 1. the usage of PIDs in data infrastructures, 2. the interoperability of data infrastructures through harmonized interfaces and APIs, and 3. the implementation of common semantic concepts. Other topics will emerge as required by the participants. Other information sources will deal with community recommendations and the establishment of a knowledge base to cover some general information.

The portal's design will be updated according to the needs of the community as it is developed. We hope that the F-IMP and the portal will support the harmonization and improvement of data management processes towards the further implementation of a FAIR data space.

Please assign your poster to one of the following keywords.

Tools

In addition please add keywords.

Strategy, Community Dialogue, Recommendations

Please assign yourself (presenting author) to one of the stakeholders.

other (please specify)

Please specify "other" (stakeholder)

Data Infrastructure Provider, Data Curators

Primary authors: PÖRSCH, Andrea (HMC Hub EE at GFZ); SÖDING, Emanuel (GEOMAR); BUTTIGIEG, Pier Luigi (GEOMAR Helmholtz-Zentrum für Ozeanforschung, Kiel, Germany); WEINELT, Martin (GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel); LORENZ, Sören (GEOMAR Helmholtz Centre for Ocean Research Kiel)

Presenters: PÖRSCH, Andrea (HMC Hub EE at GFZ); SÖDING, Emanuel (GEOMAR)

Session Classification: Postersession I

Track Classification: Postersession

Contribution ID: **94** Type: **not specified**

Join the Dots: creating metadata for a collection of 80 million items at The Natural History Museum, London

Thursday 6 October 2022 12:30 (45 minutes)

Details of less than 10% of the 80 million individual items in the collection at the Natural History Museum can be obtained via our Data Portal but much of it remains undigitized with other data associated with the collection recorded but not delivered in a coherent system. In 2018, 77 staff at the Natural History Museum, London, took part in a successful collections assessment exercise. 17 questions provided details of the Condition, Importance/Significance, Information available and Outreach use/potential about 2,602 Collection Units covering the entire Natural History science departments and Library and Archives. Results can be displayed and filtered via a bespoke dashboard in Microsoft Power BI, accessed via a web link available internally to all staff. The project successfully recorded the expertise of the curatorial staff and produced the first comprehensive assessment of the Natural History Museum's collection. The methodology includes first attempts to automate scoring to include other data from our collections management systems such as environmental conditions and completeness of data coverage for the individual items that we deliver via our data portal. A few case studies are provided here to show how we have used this data and continue to refine the process of data capture and delivery for analysis with a key example showing how we are using this data to plan a major move for 40% of our collection.

Primary author: Dr MILLER, Giles (The Natural History Museum London)

Presenter: Dr MILLER, Giles (The Natural History Museum London)

Session Classification: Keynote II