Host Lab Report 2020
Karlsruhe Institute of Technology

Helmholtz Metadata Collaboration (HMC) Platform

Wolfgang Süß, Institute for Automation and Applied Informatics (IAI)
Thomas Jejkal, Rainer Stotzka, Steinbuch Centre for Computing (SCC)
Version June 11, 2021

2020

<HMC>

Contents

Achievements and Highlights

In 2019 and 2020 the operation of the Helmholtz Metadata Collaboration Platform has been established at Karlsruhe Institute of Technology (KIT). The Institute for Automation and Applied Informatics (IAI), conducting research and development for sustainable energy systems and complex industrial and laboratory processes, is responsible for Metadata Hub Energy in WP 3. The Steinbuch Centre for Computing (SCC) with its competences in IT, High Performance Computing and Information & Data Science conducts the work in WP 2 FAIR Data Commons: Overarching Technical Services.
Both institutions have implemented the working groups to fulfill the HMC tasks. They are closely cooperating and fortnightly meetings ensure the general information exchange and discussions on novel developments on working level.
Close cooperation has been also incorporated with all other HMC working groups in the Helmholtz Association, with the Research Fields Energy, Information & Data Science, and with the Research Data Management Team at KIT.

WP 2.1 HMC FAIR Data Commons: Overarching Technical Services

The main objective of WP 2 is the implementation and facilitation of generic technical services supporting discoverability, accessibility, interoperability, and reusability (FAIR) of research data using metadata for all research fields in the Helmholtz association.

In WP 2.1 the implementation and facilitation of the FAIR Digital Object (FAIR DO) Model enable novel methods to preserve, manage, analyze and exploit cross-disciplinary big research data collections. FAIR DOs describe a concept of virtual data objects that has been developed and used by the Research Data Alliance (RDA, http://rd-alliance.org) in various Working and Interest Groups. FAIR DOs represent data, software or other research resources. They are uniquely identified by Persistent Identifiers (PID) and metadata rich enough to enable them to be reliably found, used and cited.
FAIR DOs are currently implemented world-wide, in many RDA groups, in the European Open Science Cloud (EOSC), as well as in other initiatives with the need to design and develop large integrative research data infrastructures. They have been recognized as essential components in the research data ecosystem to facilitate FAIR data (EOSC Interoperability Framework, https://www.eoscsecretariat.eu/sites/default/files/eosc-interoperability-framework-v1.0.pdf) and are also adopted in the German National Research Data Infrastructure NFDI.

Scientific Products

Regarding the creation of necessary base services, a realization roadmap has been defined based on the scientific requirements, existing expertise, and state of the art as well as identified gaps. Following this roadmap, WP 2.1 was able to achieve essential milestones towards the first implementation of FAIR Digital Objects and make them available in the form of initial services:

Service (Version)	Short Description	Source Code/Service URL
Collection Registry (1.1.0)	Creation and management of collections of digital resources. It allows to create virtual collections of heterogeneous research data located in distributed data resources.	https://github.com/kit-data-manager/collection-api
Metastore (0.2.4)	The metadata storage system consisting of metadata schema registry and metadata repository allows to manage huge collections of metadata documents.	https://github.com/kit-data-manager/metastore2
Metadata Standards Catalog (1.0)	Metadata catalog for registering and discovering existing metadata standards, see text below.	https://msc.datamanager.kit.edu/
Web Annotation Protocol Server (1.0.0)	Implementation of the W3C Web Annotation Data Protocol standard for preserving and managing annotations of research data.	https://github.com/kit-data-manager/wap-server

In addition, other services are in an advanced stage of implementation, such as a service for facilitating the easy creation of PIDs, for building search indices or for data versioning using an enhanced storage concept saving disk space. All services are currently being integrated in a testbed, which will be made available to all Metadata Hubs for demonstration purposes soon.
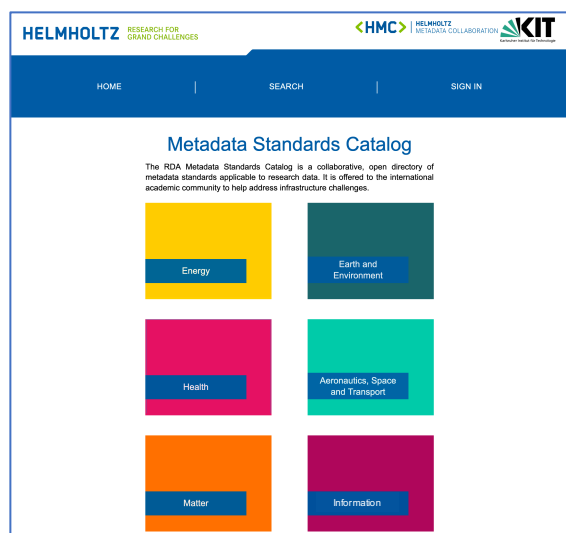

Fig.: Web interface of the Metadata Standards Catalog https://msc.datamanager.kit.edu/ .

The HMC Metadata Standards Catalog (https://msc.datamanager.kit.edu/) is a web service that will be provided for all Helmholtz centers as well as the NFDI. The catalog provides a list of metadata standards and supports scientists to identify standards that meet their requirements. Adopted from the Research Data Alliance, the catalog offers commonly used standards that are sorted according to the research fields of the Helmholtz Association. Furthermore, the catalog can be easily extended for sharing new schemas, standards, and tools used by Helmholtz scientists. End of 2020 the development of the catalog has been completed; in early 2021 the web service will be openly available.

Raised Third-party Funds and Cooperations

NFDI

In the context of National Research Data Infrastructures (NFDI), WP 2.1 representatives are involved in a couple of consortia which are benefiting from the experience gained and from the upcoming service portfolio of HMC.

HMC WP 2.1 is involved in two successful consortia of the first round that have started their work in October 2020: NFDI4Chem and NFDI4Ing. We are contributing to metadata and terminology services, provisioning of a metadata hub, research data repositories, electronic lab books, research data storage, and the overall infrastructure and software architecture.

In the current NFDI call WP 2.1 has contributed to NFDI-MatWerk, NFDI4Earth and PUNCH4NFDI with components building up the distributed ecosystem of metadata services. In all consortia the results of the FAIR Digital Objects will be introduced.

DFG Collaborative Research Center "Episteme in Motion"

CRC 980 "Episteme in Motion" has started its third funding phase in summer 2020. The reviewers "were impressed" by the infrastructure project INF "Corpora in Motion" (http://www.sfb-episteme.de/en/teilprojekte/informationsinfrastruktur/index.html) with research data repositories, Metastore and vocabulary services that are in development of WP 2.1.

Horizon 2020 project "NFFA Europe Pilot"

The European project (https://www.nffa.eu) has been successfully evaluated and will start in March 2021. The HMC metadata services will play a central role in the research data management of a consortium of 23 partners in the field of "Nanoscience Foundries and Fine Analysis".

Other

As early adopters of the technical services two pilot projects in HMC are supported:

- Metadata Repository instance for a photovoltaic system with battery storage

  This cooperation with the Metadata Hub Energy at IAI will adopt the Metastore. In a later phase, also PID support will be integrated in order to identify energy measurement devices, like Smart Meters, persistently and globally unique.

- Reproducibility of simulation workflows

  In this pilot in cooperation with the Metadata Hub Information at Forschungszentrum Jülich a metadata repository based on Metastore for storing provenance metadata as well as tools for automatic provenance tracking for materials simulation workflows will be realized.

In the research field Information & Data Science HMC WP 2.1 will contribute in the Joint Labs / Cross-Cutting Activities

- Virtual Materials Design (VMD) with ontologies for HMC simulations and
- Integrated Model and Data Driven Material Characterization (MDMC) with Metastore and vocabularies for materials research.

Support Activities

In addition to the work on the basic services realizing the FAIR Digital Object Model, WP 2.1 was active in the area of consulting and support.
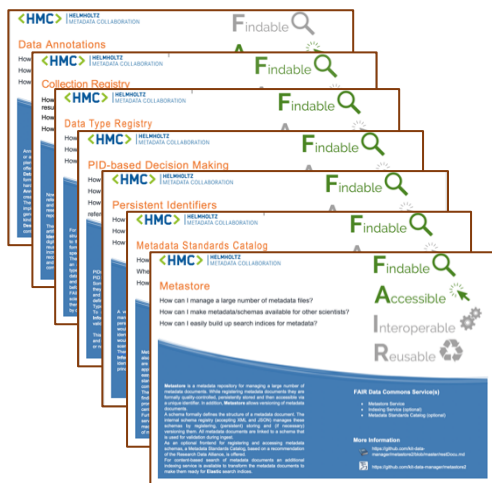

Fig.: Service flyers introducing HMC service components.

WP 2.1 provided documentation (e.g. see service flyers https://helmholtz-metadaten.de/en/fair-data-commons/use-cases, left), advice and support at many points in the national scientific community. Support was provided to various consortia of the NFDI, HMC projects, Metadata Hubs, the Data Hub in Earth & Environment, and scientists from other research fields.

Especially consortia that planned to propose projects for the HMC Project call required assistance concerning the HMC goals and upcoming software and services, e.g. Metastore and Metadata Standards Catalog.

In 2020, representatives from WP 2.1 took over the leadership of the Research Data Alliance (RDA, https://rd-alliance.org/) "IG Data Fabric" providing an international platform for sharing experiences and implementations to harmonize the use of FAIR Digital Objects (https://www.rd-alliance.org/group/data-fabric-ig/wiki/rda-ig-data-fabric-fair-digital-objects). Within the RDA IG Data Fabric, two sessions at virtual RDA Plenaries as well as monthly virtual meetings were organized and conducted getting much international attention and contributions.

Fig.: The objectives of the RDA IG Data Fabric complement the activities in HMC WP 2.1.

Events:

- Jan. 15, Internal Kickoff Meeting of Office, Metadata Hub Coordinators and WP 2, Heidelberg
- Mar. 2-3, HMC All Hands Meeting, Berlin
- Mar. 18-20, Virtual RDA Plenary 15, organization of a session (https://www.rd-alliance.org/plenaries/rda-15th-plenary-meeting-australia)
- Jul. 02-03 NFDI Metadata Workshop, organization and moderation (https://nfdi4ing.de/gone-virtual-two-day-nfdi-metadata-workshop-on-july-2-3rd-2020/)
- Sep. 23, Helmholtz Virtual Data Science Career Day (https://www.helmholtz-hida.de/angebote/veranstaltungen/detail/helmholtz-virtual-data-science-career-day/)
- Oct. 22-23, CODATA and GO FAIR, International FAIR Convergence Symposium (https://www.go-fair.org/events/international-fair-convergence-symposium/)
- Nov. 03-06, HMC Retreat

- Nov. 09-12, Virtual RDA Plenary 16, organization of a session (https://www.rd-alliance.org/plenaries/rda-16th-plenary-meeting-costa-rica-virtual)
- Nov. 19, Presentation of Metadata & HMC, Data Hub Earth & Environment
- Dec. 10, Metadata Discourse
- 4 Presentations of HMC to NFDI consortia and the NFDI Directorate

WP 3 HMC – Metadata Hub Energy

The centers KIT, HZDR and DLR are involved in the Metadata Hub Energy: KIT and HZDR through direct participation, DLR through a self-financed part. A special feature of the research field Energy is that the standardization of research data is not as advanced as in other research areas. This is a great challenge for the work in the Metadata Hub Energy. The tasks in WP3 are divided into three sub-items. The work on this is described below.

WP 3.1 Coordination and Management

The Energy Hub was implemented for the energy research field and is functional. The main task of WP 3.1 is to build a metadata community and to network with other committees. In the first place, the cooperation between the partners HZDR and DLR was established and expanded. Then partners from other existing projects were involved. HMC has also been presented on several occasions.

The following list shows the activities in WP 3.1 in detail.

WP 3.1.1 Inclusion of community expertise
- Regular meetings with HZDR and DLR
- Crosslinking with the Energy Lab 2.0 (https://www.elab2.kit.edu/index.php) project. Partners are FZJ, DLR and KIT
- Participation in ES2050 (Energie System 2050, https://www.helmholtz.de/forschung/energie/energie-system-2050/)
- Collaboration in the various working groups within HMC

WP 3.1.2 Development and support of a 'metadata community' in the research area
- Work has begun on building and maintaining the metadata community in the energy research field
- Presentation of HMC
    - in the LLEC project (Living Lab Energy Campus, https://www.fz-juelich.de/llec),
    - in the StoRIEs meeting (EU project proposal - Storage Research Infrastructure Ecosystem,
    - at DLR,
    - in the EERAdata project (European Energy Research Alliance, https://www.eeradata.eu/),
    - at Helmholtz AI Kick-off meeting.
- Contacts to Open Energy Platform (https://openenergy-platform.org/)

WP 3.1.3 International networking, committee work, harmonization and standardization
- Cooperation with EERAdata (European Energy Research Alliance, https://www.eeradata.eu/)
- Participation in the working group AK 952.0.10 of the DKE (Deutsche Kommission Elektrotechnik Elektronik Informationstechnik in DIN und VDE, https://www.dke.de). The goal is the further development of the IEC 61850 standard.

WP 3.2 Knowledge and Communication

One task in this work package is the collection of metadata standards, vocabularies and ontologies. Another important task is the mapping of the research data landscape in the energy research area. Significant contributions were made to both tasks in the Metadata Hub Energy.

Details are shown in the following list:

WP 3.2.1 Development and provision of an information base on metadata (vocabularies), ontologies and standards

- An extensive collection of metadata standards, vocabularies, and ontologies has been started and documented in the Hub internal wiki.

WP 3.2.2 Research data expertise landscape

- A catalogue of questions for mapping the research data landscape was created and an online survey was generated from this. This online survey was tested internally and further developed. Our experience from this was also brought into the Task Force Survey of HMC.
- The mapping of the research data landscape started with the Energy Lab 2.0 project (https://www.elab2.kit.edu/). The developed survey was used for this purpose.

WP 3.2.4 Consulting to centers, projects and researchers

- Start of the consultation in the Energy Lab 2.0

WP 3.3 Components and Processes

A pilot project was defined in this work package and the implementation of the project started. The pilot project is a typical example from the energy research area showing how the storage of data according to the FAIR principles can look like in practice. Existing metadata standards and ontologies were evaluated and adapted for this project.

For the implementation, tools from WP 2.1 are used and necessary tools are developed in-house.

More details are in the list below:

WP 3.3.1 Establish processes, tools and services for the development of research data collections

- In order to demonstrate processes, tools and services for the exploitation of research data, a pilot project has been started. The objective of the project is to model an existing photovoltaic (PV) system with battery storage by using metadata standards and ontologies. Therefore, metadata standards (e.g. IEC61850) and ontologies (e.g. CIM, PV-TONS) were searched and evaluated. The PV system model as well as its measurement data are used to demonstrate the FAIR principles for a continuous data producer (comparable to Energy Lab 2.0 facilities).
- Exchange with Metadata Hub Matter on the interdisciplinary challenges of the pilot project.

WP 3.3.3 Tools for handling various metadata standards

- The IEC61850 standard was mapped to a JSON representation and an own ontology was developed for the PV system description (since a usable ontology could not be found).
- Tools like Metastore from WP2.1 are used to handle metadata descriptions.

KIT – Host Lab Key Performance Indicators

	KPI	Indicator	Value
P1	Number of mapped Metadata Schemas, Processes and Data Sources	Count any schema, instrument, lab, process mapped during the mapping process.	107
P2	Number of workshops organized	Aggregate number of total courses and workshops conducted.	4
P3	Number of Helmholtz external institutions working with HMC on metadata topics	Count partners actively working with HMC on Metadata integration (e.g. NFDI, universities etc.)	10 (~75)
P4	Number of HMC Presentations	Number of presentations at conferences or self-organized promotion events.	22

Contributions to the KPI are:

WP 2.1 HMC FAIR Data Commons

P1:   100 registered and linked metadata schemas and standards in the Metadata Standards Catalog,
        (3 additional metadata schemas to administrate Metastore, collections and the Metadata Standards Catalog),
        1 W3C Web Annotation Data Model

P2:   1 Metadata discourse,
        2 RDA sessions,
        1 NFDI Metadata Workshop

P3:   5 NFDI consortia (~50 institutions outside the Helmholtz Association),
        1 CRC (2 institutions outside the Helmholtz Association),
        1 H2020 Project (20 partners outside the Helmholtz Association),
        (RDA members cooperating in the groups were not taken into account)

P4:   2 Research Data Alliance IG Data Fabric,
        1 Research Data Alliance IG Repository Systems 4 Research Data,
        1 Data Hub Earth & Environment,
        2 HMC Retreat,
        1 Metadata discourse,
        4 Presentations of HMC to NFDI consortia and the NFDI Directorate,
        1 Helmholtz Virtual Data Science Career Day,
        5 Presentations in KIT boards

WP 3 Metadata Hub Energy

P1:   6 Ontologies and Standards (IEC 61850, CIM (IEC 61970), PV-TONS, SEAS-Photovoltaic, SoftCost, OpenEnergy Ontology)

P3:   3 External institutions (EERAdata, Open Energy Platform, LLEC)

P4:   5 Presentations (EERAdata, LLEC, StoRIEs, DLR, HelmholtzAI – Kick Off Meeting)

Operational Adjustments

KIT contributes also significantly to other working packages in HMC:

- "WP 2.2 Components and Processes – Realization of the FAIR Principles" with tools and software components,
- "Metadata Hub Information" in the Joint Labs / Cross-Cutting Activities,
- "WP 1.1.4 International networking, boards, harmonization, and standardization" by their engagement in the NFDI and the Research Data Alliance, e.g. elected member of the Technical Advisory Board and co-chairing RDA Interest Groups,
- "WP 1.1.2 Communication and Public Relations" by creating web content, and
- "WP 1.2 Teaching and Training"