Modern Earth system science visualization and exploration techniques

ESM Data Exploration with the Model Data Explorer

ESSI4.1 — EGU, Vienna, Austria — April 29th, 2023

Talk

Session Intro

2-Minute Madness

General

Author

Funding

Abstract

Implementation

Core

Plugins

Federation

Maintenance

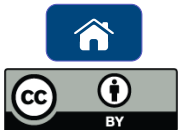
Resources

DOI 10.5194/egusphere-egu23-3624

How to navigate

This presentation has been prepared for a PICO presentation at the EGU 2023 in Vienna, Austria. To facilitate the navigation, a lot of hyperlinks are used. Almost every item in this presentation is clickable:

- click buttons like this to be linked to other connected frames
- click the navigation bar above with the sections, Home, Help, etc. (including the dots) to navigate in the presentation
- click on navigation buttons like this < 1/2 > to show you more of the current frame.
- click on many of the images to get more information or a close-up
- click the  icon to go back to the menu
- click the buttons at the lower left and lower right that bring you to the next slide

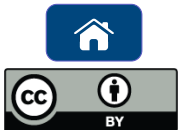


Session title page ▶

How to navigate

This presentation has been prepared for a PICO presentation at the EGU 2023 in Vienna, Austria. To facilitate the navigation, a lot of hyperlinks are used. Almost every item in this presentation is clickable:

- click buttons like this to be linked to other connected frames
- click the navigation bar above with the sections, Home, Help, etc. (including the dots) to navigate in the presentation
- click on navigation buttons like this < 2/2 > to show you more of the current frame. For example this text.
- click on many of the images to get more information or a close-up
- click the  icon to go back to the menu
- click the buttons at the lower left and lower right that bring you to the next slide



Session title page ▶

Modern Earth system science visualization and exploration techniques

The balancing act between complex information, broad functionality and simple illustration

ESSI4.1 — EGU, Vienna, Austria — April 29th, 2023

Conveners:
Tobias Kerzenmacher, Christof Lorenz, Ugur Cayoglu, **Philipp S. Sommer**

Karlsruher Institut für Technologie

DataHub
Research Field Earth & Environment

Helmholtz-Zentrum hereon

◄ Help

Challenges and opport . . . ►

Challenges and opportunities

Earth system science data are getting increasingly important

- as decision support for stakeholders
- for other end users far beyond the scientific domains
- within project collaborations
- Institutes, Groups and projects are reviewed with respect to the data products they publish

But higher temporal and spatial resolutions of modeling and remote sensing approaches lead to ever-increasing data complexity and volumes

Challenges and opportunities

Earth system science data are getting increasingly important

- as decision support for stakeholders
- for other end users far beyond the scientific domains
- within project collaborations
- Institutes, Groups and projects are reviewed with respect to the data products they publish

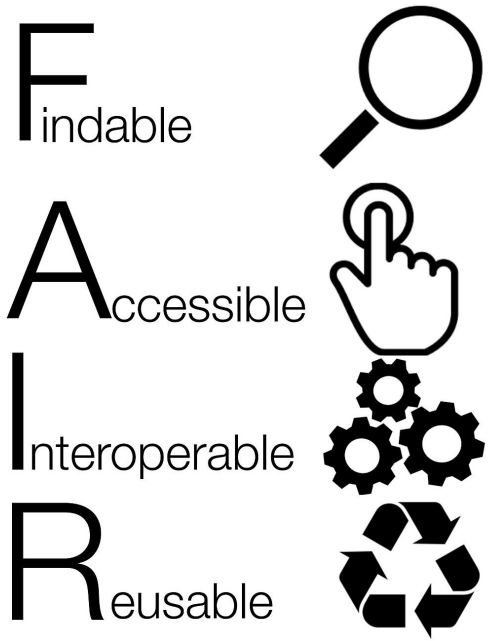
But higher temporal and spatial resolutions of modeling and remote sensing approaches lead to ever-increasing data complexity and volumes

We want to prevent

- data that cannot be found
- outdated and insecure software
- always reinventing the wheel

Data standards

Making data available can push the use of data standards that improve data findability, accessibility, interoperability and reusability (FAIR).

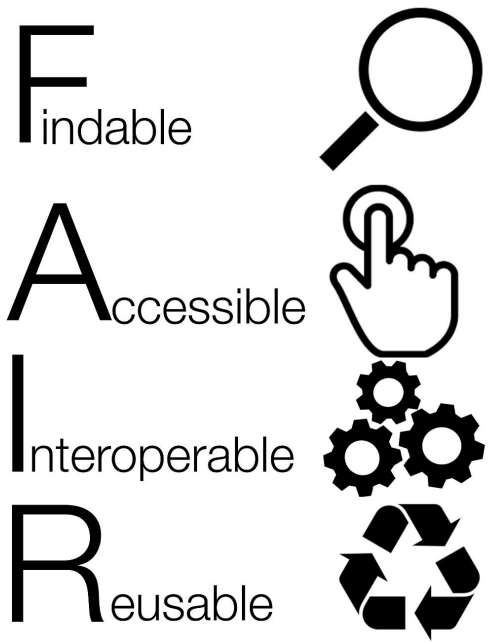


Data standards

Making data available can push the use of data standards that improve data findability, accessibility, interoperability and reusability (FAIR).

Tool tips to improve data management

- use OGC standards for visualization
- do not only visualize, provide download as well
- correctly cite the data and publications you are showing
- ask supplied data to fulfill certain metadata standards
- provide instructions (e.g. in form of Standard Operating Procedures) on how to get different data into your software



Research Software Recommendations

Tips for improving persistence

- make your software modular
- invest time in the conceptualization of a software framework
- make it open-source
- make it available in an open gitlab or on github
- make your software installable in different infrastructures
- package your software
- version your software
- containerize your software deployment-ready
- provide login via eduGAIN or other AAI

Research Software Recommendations

Tips for improving persistence

- make your software modular
- invest time in the conceptualization of a software framework
- make it open-source
- make it available in an open gitlab or on github
- make your software installable in different infrastructures
- package your software
- version your software
- containerize your software deployment-ready
- provide login via eduGAIN or other AAI

Tips for implementing tests

- invest time in unit tests
- invest time in stakeholder tests
- automate your documentation
- use continuous integration for testing your software

Research Software Recommendations

Tips for improving persistence

- make your software modular
- invest time in the conceptualization of a software framework
- make it open-source
- make it available in an open gitlab or on github
- make your software installable in different infrastructures
- package your software
- version your software
- containerize your software deployment-ready
- provide login via eduGAIN or other AAI

Tips for implementing tests

- invest time in unit tests
- invest time in stakeholder tests
- automate your documentation
- use continuous integration for testing your software

Tips for documentation

- invest time in API documentation
- describe your framework
- provide detailed installation instructions

Our session

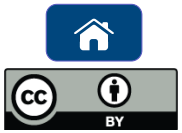
A diverse set of tools that make Earth System Science Data available on the web

- Generic data exploration frameworks
- Personalized data exploration frameworks driven by a specific scientific use-case
- Transfer- and outreach-driven exploration frameworks

Our hope

Establish a transdisciplinary community of scientists, software-developers and other experts in the field of data visualization in order to exchange and to give a state-of-the-art overview of tools, interfaces and best-practices.

◀ Research Software Re . . .



MDE title page ▶

Model Data Explorer

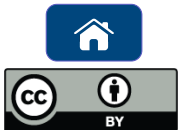
ESM Data Exploration with the Model Data Explorer

EGU, Vienna, Austria, April 29th, 2023

Philipp S. Sommer

Helmholtz Coastal Data Center
Helmholtz-Zentrum Hereon

◀ Our session

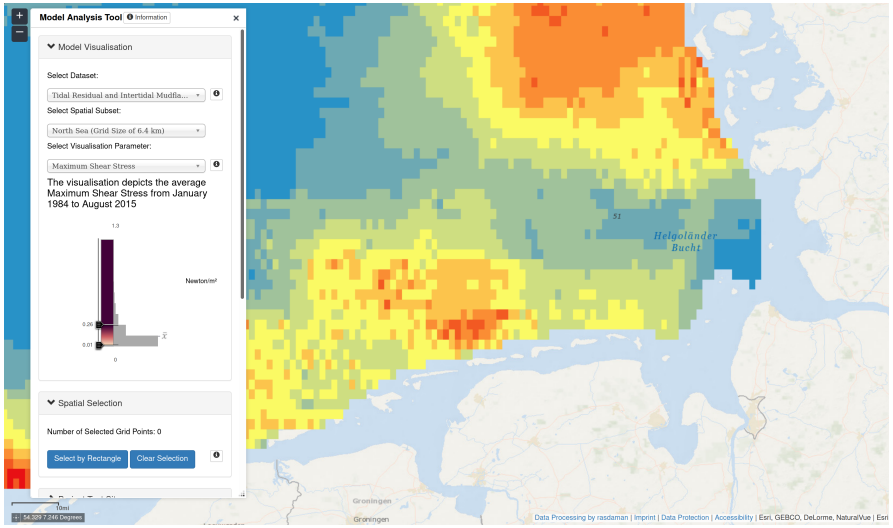


A central platform to a . . . ▶

A central platform to access Model Data

Core features

- View 4D Model-Data on a map
- Categorized: One page per project, institution, author, etc. with all corresponding model runs
- Search and filter model runs and groups
- Compute and compare statistics on/of subsets of Model Data
- Download the raw data for the selected region

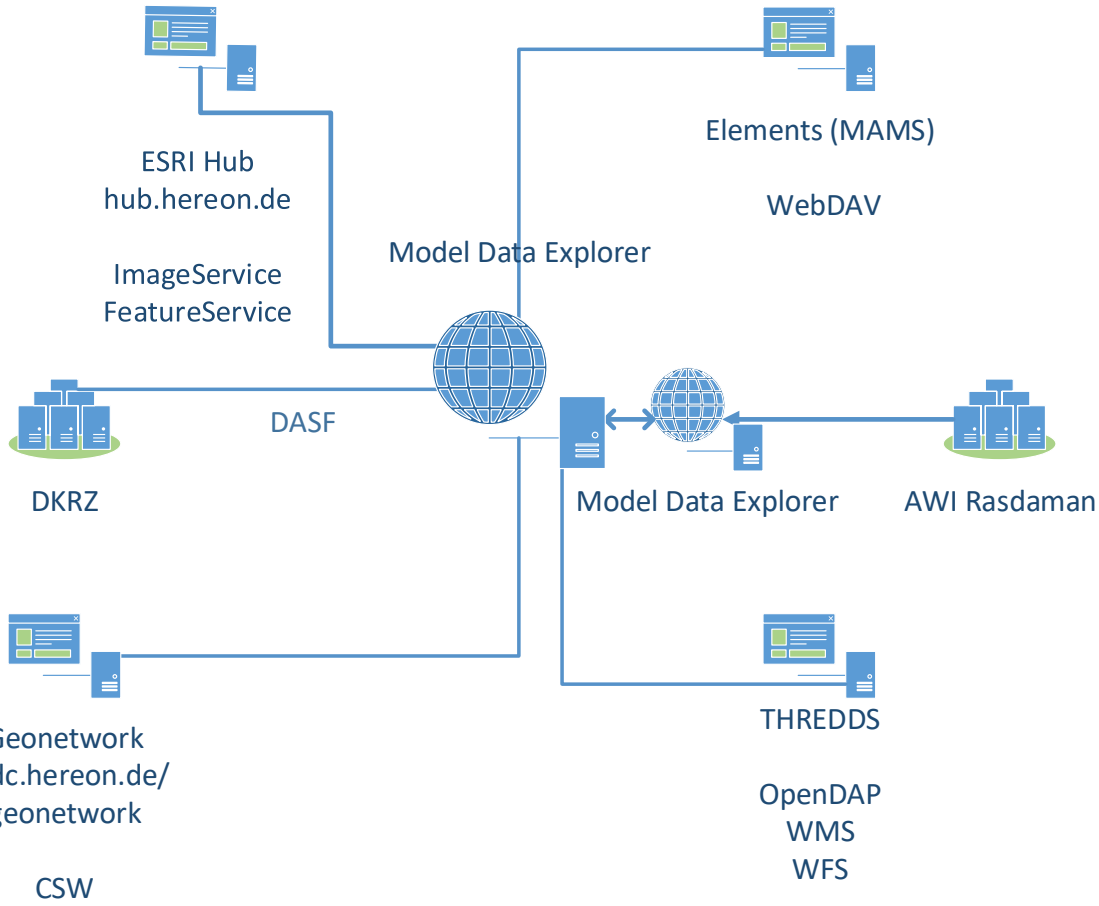


◀ MDE title page
A modular and federat . . . ▶

A modular and federated system

- Single configuration interface for multiple connected services
- multiple standard interfaces
- Federation to connect multiple research centers

More . . .



◄ A central platform to a . . .

Author ►

Author

Philipp S. Sommer

Helmholtz Coastal Data Center (HCDC)
Helmholtz-Zentrum hereon GmbH
Max-Planck-Straße 1
DE - 21502 Geesthacht

E-mail philipp.sommer@hereon.de
Telefon +49 4152 87 2126
Github https://github.com/Chilipp
Webpage https://www.philipp-s-sommer.de

Co-Authors

Helmholtz-Zentrum Hereon

- Linda Baldewein
- Hatef Takyar
- Rehan Chaudhary
- Housam Dibeh
- Max Böcke
- Ulrike Kleeberg

Karlsruhe Institute of Technology

- Mostafa Hadizadeh
- Christof Lorenz

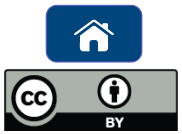
Alfred-Wegener-Institut

- Tilman Dinter
- Stefan Pinkernell

Geomar

- Klaus Getzlaff

◀ Author

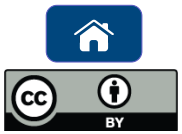


Funding ▶

Funding

The conceptualization of the Model Data Explorer has been made possible through the DataHub funding. Initial developments for the THREDDS-Server management are funded through Cat4KIT.
We hope to get funding through an NFDI4Earth Pilot (`https://www.nfdi4earth.de/`) for the core implementation of the model data explorer.

◀ Co-Authors



Abstract ▶

Abstract I

ESM Data Exploration with the Model Data Explorer

Making Earth-System-Model (ESM) Data accessible is challenging due to the large amount of data that we are facing in this realm. The upload is time-consuming, expensive, technically complex, and every institution has their own procedures.
Non-ESM experts face a lot of problems and pure data portals are hardly usable for inter- and trans-disciplinary communication of ESM data and findings, as this level of accessibility often requires specialized web or computing services.
With the Model Data Explorer, we want to simplify the generation of web services from ESM data, and we provide a framework that allows us to make the raw model data accessible to non-ESM experts.
Our decentralized framework implements the possibility for an efficient remote processing of distributed ESM data. Users interface with an intuitive map-based front-end to compute spatial or temporal aggregations, or select regions to download the data. The data generators (i.e. the scientist with access to the raw data) use a light-weight and secure python library based on the Data Analytics Software Framework (DASF, `https://digital-earth.pages.geomar.de/dasf/dasf-messaging-python`) to create a back-end module. This back-end module runs close to the data, e.g. on the HPC-resource where the data is stored. Upon request, the module generates and provides the required data for the users in the web front-end.

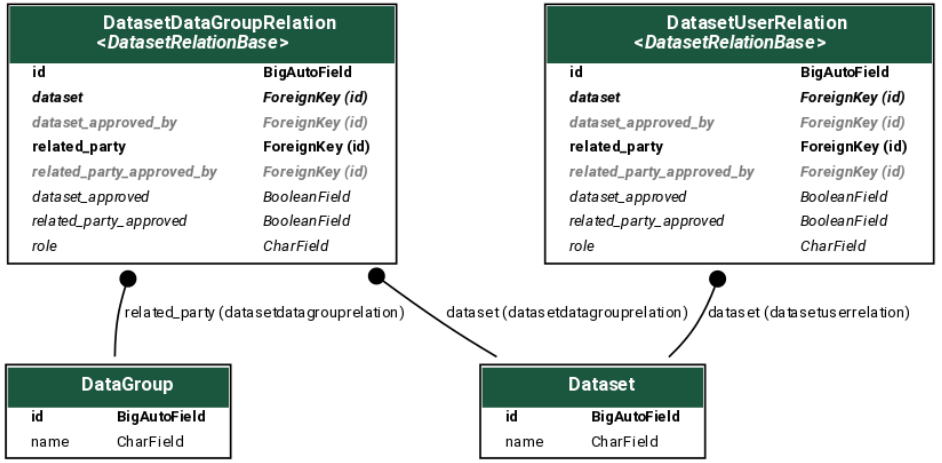
Abstract II

ESM Data Exploration with the Model Data Explorer

Our approach is intended for scientists and scientific usage! We aim for a framework where web-based communication of model-driven data science can be maintained by the scientific community. The Model Data Explorer ensures fair reward for the scientific work and adherence to the FAIR principles without too much overhead and loss in scientific accuracy.
The Model Data Explorer is in the progress of development at the Helmholtz-Zentrum Hereon, together with multiple scientific and data management partners in other German research centers. The full list of contributors is constantly updated and can be accessed at
`https://model-data-explorer.readthedocs.io`.

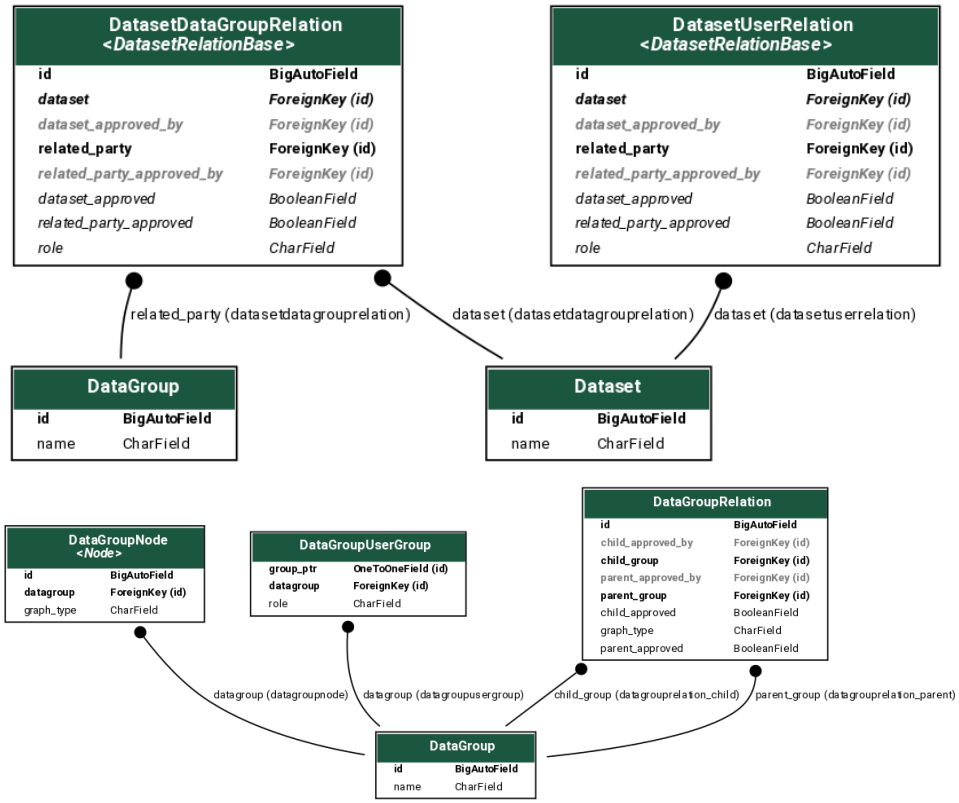
Core

The core package of the Model Data Explorer, `mde-core`, implemented with Django, implements so-called *Datasets*. A dataset might be identified with a specific experiment to run a model. It gets a unique handle and UUID and serves as a kind of namespace for the services associated with it. This makes resources related to a dataset better findeable as everything is linked to a specific dataset.



Core

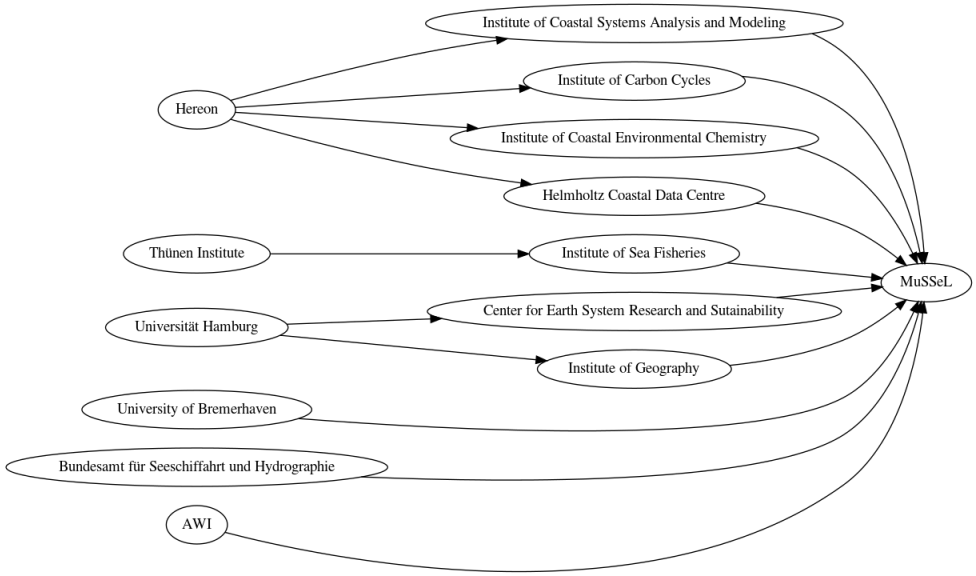
`mde-core` additionally implements data groups. This can be anything from a project to a department or an institution. DataGroups can own datasets and they will be listed on their detail page. DataGroups can also own other DataGroups and as such inherit ownerships on the datasets. `mde-core` uses a directed acyclic graph implementation (OmenApps, 2021) to efficiently manage data group relations and permissions.



DataGroup relations

In Research

- Many partners are involved in different projects
- Data managers need to be able to configure metadata and create services
- Relations should be built from the metadata and vice-versa



In the Model Data Explorer

- Relationships between groups are represented with a directed acyclic graph
- Datasets are shown on each group site
- Each institute, department, unit, project, etc. gets its own page
- Maintainability is guaranteed through data group relations

◀ Core

Plugins ▶

Plugins

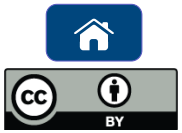
Each research center has a different IT infrastructure and works with different frameworks. The aim of the model data explorer is to integrate these different systems into a user-friendly web-application. In order to be able to deploy the model data explorer at the different research centers, we aim for a modular structure where each component can be added or removed depending on the setup at the hosting institution. This modular structure is implemented via plugins. Each plugin is a python package containing a django app that can be added or removed from the configuration. We distinguish between three plugin categories: core, framework and service plugins.

Framework plugins

Core

Service plugins

◄ DataGroup relations



Framework plugins ►

Framework plugins

Framework plugins contain optionally but commonly used features for service plugins and are independent of the IT infrastructure.

mde-config

- defines a viewset-based structure where service plugins can provide configuration interfaces for individual datasets.
- provides user configuration interfaces for the `mde-core` models.

mde-federation

- combines multiple MDE instances at different institutions.
- synchronizes datasets and data groups between independant instances

mde-notifications

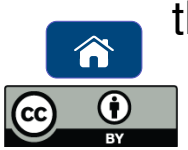
- implements the functionalities to send out notifications to the users (e.g. for confirmations required by dataset or data group owners)
- provides an interface for the user for configuring notifications

mde-oauth

- implements the possibility for the model data explorer to serve as an OAuth or SAML Provider
- gives attached services the possibility to authenticate against the MDE and get the users groups.

Read more ...

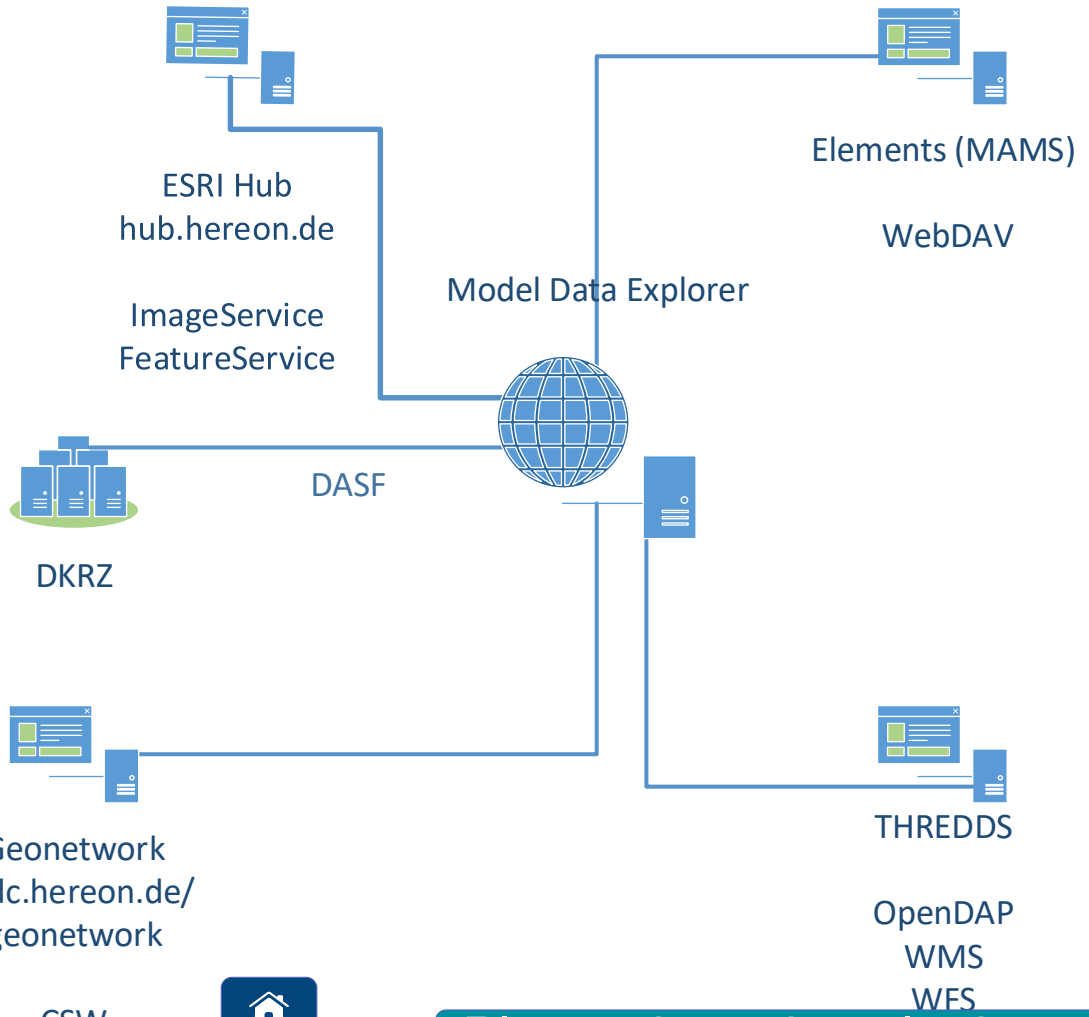
◀ Plugins



Service plugins ▶

Service plugins

Service plugins are based on the available IT infrastructure at a research institution. They connect a dataset of the `mde-core` to an external service. This way, the MDE is configuring this external service.
Service plugins should be independent of other service plugins (if possible) but they can depend on one of the framework plugins.



◄ Framework plugins

Planned service plugins ►

Planned service plugins

mde-viewer

A map-based viewer frontend for a dataset based on ArcGIS4Javascript and angular. It displays all metadata and services for an individual dataset. Services that are displayed in this viewer are configured through the mde-esri, mde-thredds, mde-analysis and mde-download plugins.

mde-thredds

This app creates catalog and configuration files for a THREDDS-Server (Caron et al., 1997) and restarts it when the configuration changed.

mde-mams

At Hereon, AWI and Geomar, we host instances of the *Media Assets Management System*, also called MAMS. Through `mde-mams`, users can request storage in MAMS to upload netCDF files. These workspaces are additionally mounted on the THREDDS-Server.

mde-csw

The `mde-csw` plugin can combine multiple ISO-files (e.g. generated via THREDDS) and provides an OGC-compliant CSW endpoint via pycsw (Kralidis et al., 2023).

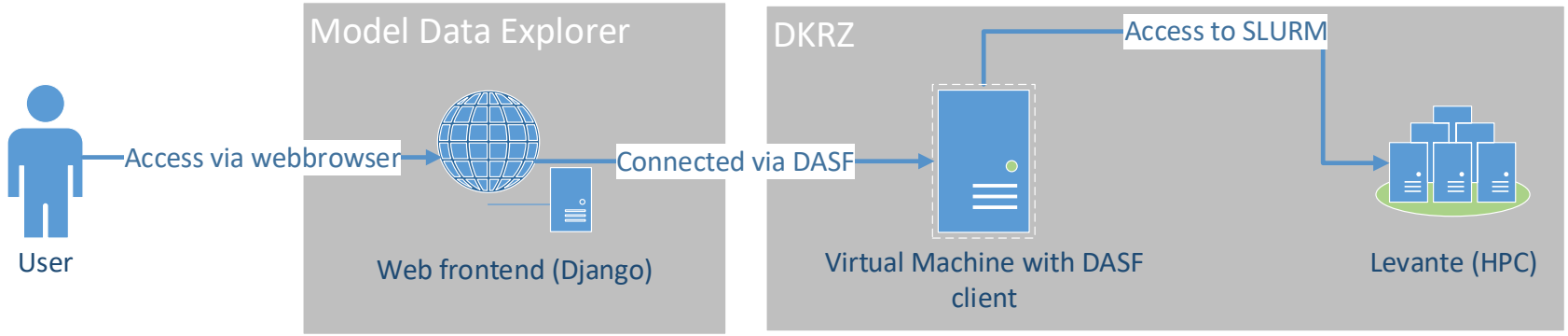
Planned service plugins

mde-analysis

The Analysis framework for the model data explorer. Based upon the Data Analytics Software Framework (DASF, Eggert et al., 2022) this plugin provides the framework to compute aggregated analysis on datasets without the need to download the raw data.

mde-download

The download framework for the model data explorer. As `mde-analysis` based upon the DASF. This plugin provides the framework to download subsets of the data.

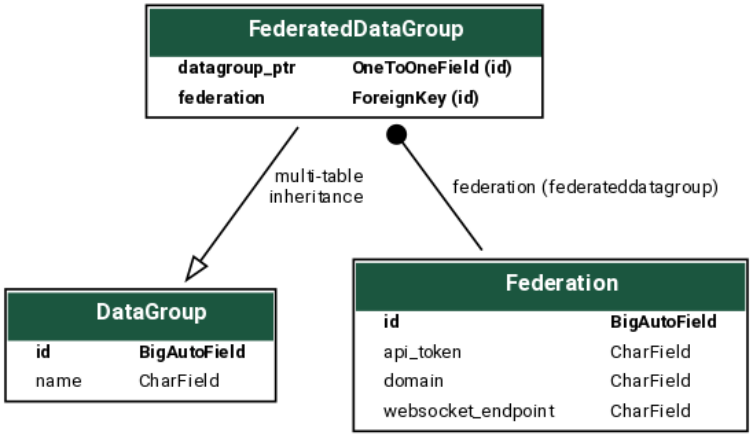


Federation

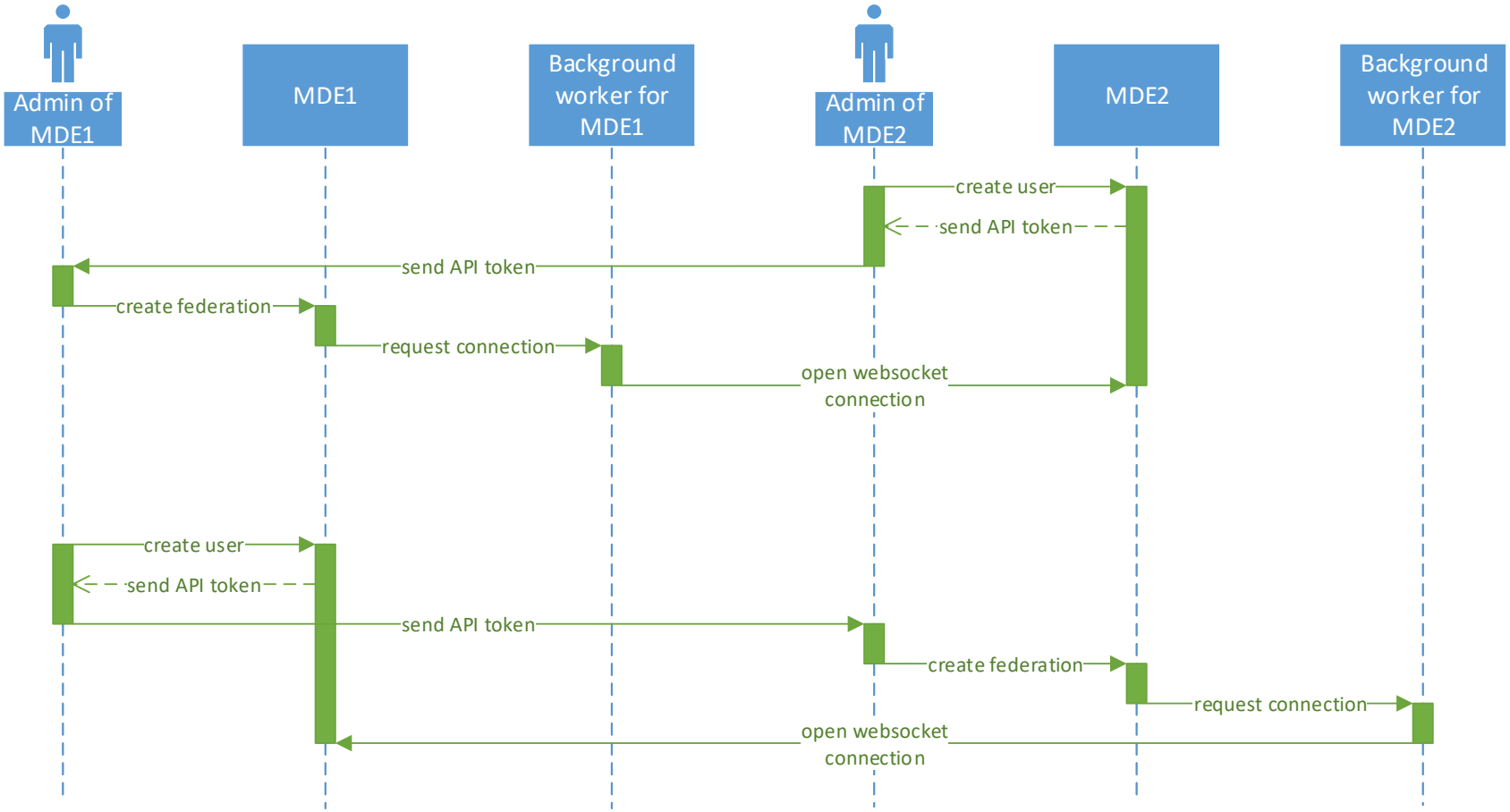
A core concept within the model data explorer is the federation between multiple MDE instances. This is essential as multiple institutions might offer different services for one single dataset. One institution might make the data available via THREDDS, the other one might make it available as ESRI `ImageService`.

Core concepts

1. Each data groups has a unique hosting MDE
2. OAuth for synchronizing user permissions between two MDE instances
3. A websocket messaging for synchronizing MDE instances



Websocket implementation for federation

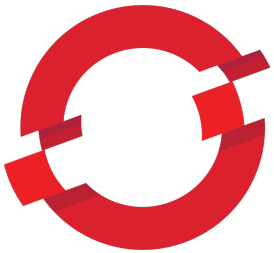


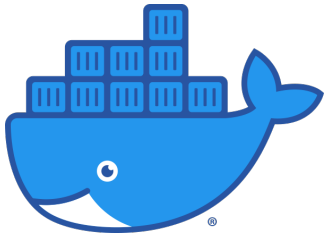
◀ Federation
Maintenance ▶

Maintenance

Templating, Packaging and Containerization

We plan to systemize and guarantee maintenance through a sophisticated templating mechanism for generating and updating plugins and docker-based deployments with cruft and cookiecutter. Standardized and flexible deployments will be implemented with helm charts for kubernetes and openshift.





◀ Websocket implement . . .

Resources ▶

Resources

Documentation

https://model-data-explorer.readthedocs.io/projects/prototype/

Source Code

https://codebase.helmholtz.cloud/model-data-explorer/

Presentation material

https://github.com/Chilipp/mde-presentation-egu2023

References I

J. Caron, E. Davis, M. Hermida, D. Heimbigner, S. Arms, C. Ward-Garrison, R. May, L. Madry, R. Kambic, and H. Johnson. Unidata thredds data server. 1997. doi: 10.5065/D6N014KG.

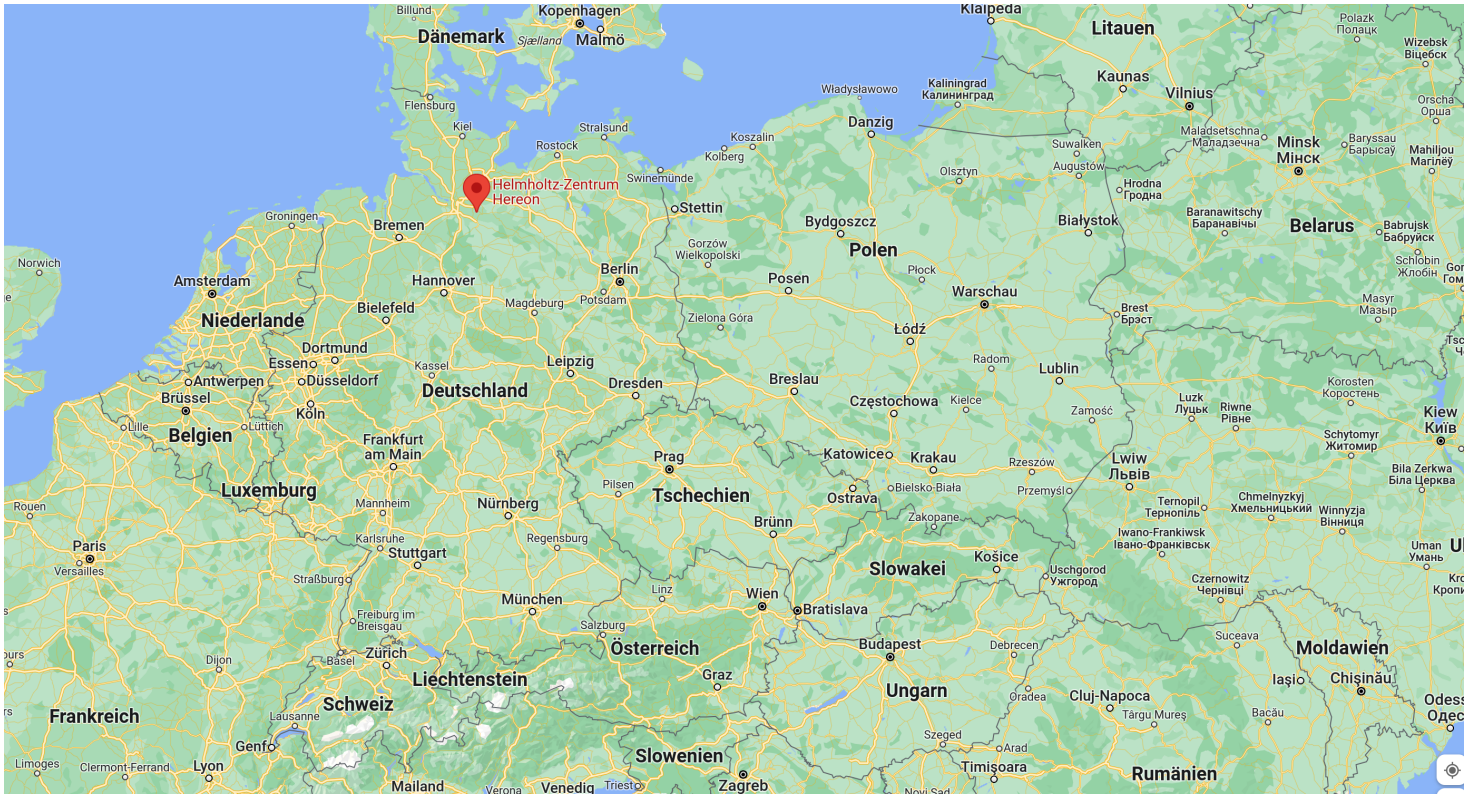
D. Eggert, M. Sips, P. S. Sommer, and D. Dransch. Dasf: A data analytics software framework for distributed environments. *Journal of Open Source Software*, 7(78): 4052, 2022. doi: 10.21105/joss.04052. URL https://doi.org/10.21105/joss.04052.

T. Kralidis, A. Tzotsos, R. G. Silva, gsaaportal, M. D. Stefano, P. van Genuchten, B. Couwenberg, KoalaGeo, A. Núñez, M. Kmet, S. Dalmasso, B. Kun, J. V. de Wauw, R. Coup, M. D. Nguyen, A. Hinz, A. Yoshiizumi, E. Rouault, K. Wilcox, A. Mercader, Filipe, A. N. Eldeen, A. Schaefer, B. Greiner, C. de Stigter, F. Frassinelli, F. Warmerdam, J. Cepicky, J. F. Dickens, and J. McKenna. geopython/pycsw: 2.6.1, Apr. 2023. URL https://doi.org/10.5281/zenodo.7839119.

OmenApps. Django & postgresql-based directed acyclic graphs, 2021. URL https://github.com/OmenApps/django-postgresql-dag/.

,

Helmholtz-Zentrum Hereon



Back to author page

Helmholtz Coastal Data Center (HCDC) - Helmholtz-Zentrum Hereon
Max-Planck-Straße 1 - DE - 21502 Geesthacht

◀ References