

# NRP and eLTER-ERIC Data Infrastructure

Tomáš Rebok

Institute of Computer Science

Masaryk University



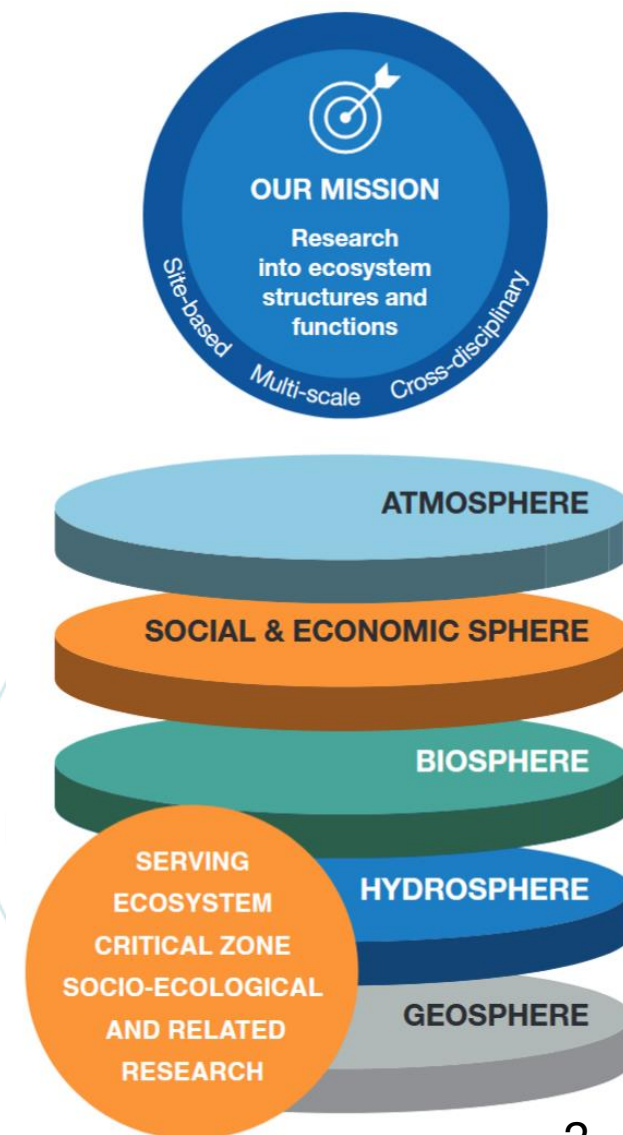
Co-funded by  
the European Union



MUNI

# eLTER Research Infrastructure (RI)

- **a pan-european RI for environmental data**
  - European network of measuring stations and long-term data collection
    - **a holistic and "site-level" approach to climate change**
    - 26 national LTER networks, ca 500 measuring stations
  - preparatory & development phase (H2020 projects)
- Partners from many European / non-European countries
  - since 2018 in ESFRI Roadmap
  - Step-1 application for eLTER ERIC in Q1/2025
- Partnerships with other European environmental research infrastructures (clustered under ENVRI-FAIR)
  - eg. ICOS, AnaEE, LifeWatch, LTER, ACTRIS, ...



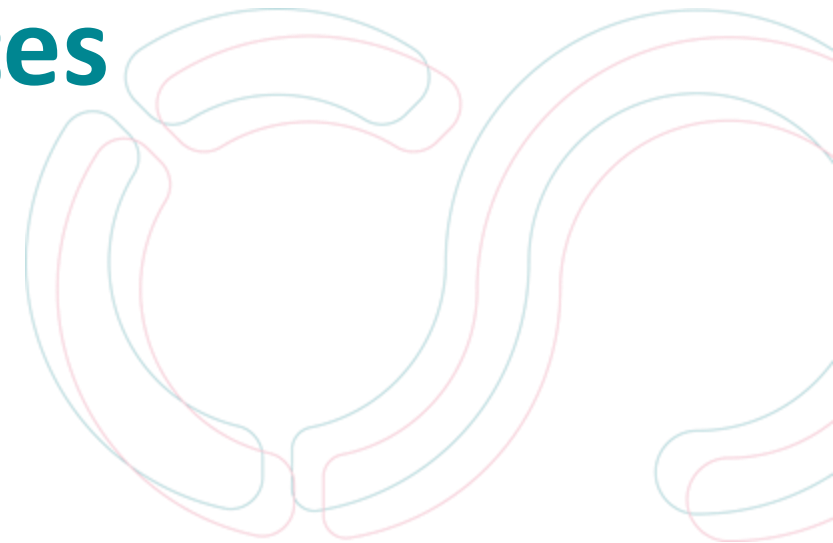
## eLTER RI – Core Services

- several central services (Topic Centres) assumed
  - monitoring and data collection, experimental infrastructure, education and training, technological innovation, and ...
- **Data Infrastructure** (eLTER CyberInfrastructure, eLTER CI)
  - **access to long-term data and metadata**
    - online/offline data ingests and collections from a network of measuring stations
  - **environment for analysis and modelling of ecosystem processes**
    - ie. data analysis environment
  - **environment for presentation of outputs (web portals, analyses, ...)**
    - ie. to publish eLTER products

# eLTER in 2023: UK CEH withdrawal

- UK Centre for Ecology & Hydrology decided to withdraw their intent in eLTER CI development
  - => the search for new partners begun
- after several meetings, **MUNI has been accepted as a new eLTER's partner** (SC meeting, Nov'23)
  - MUNI becomes a leading member of the team responsible for the development of the eLTER Cyberinfrastructure (CI) and Data Services
  - **April'24: MUNI formally joins the eLTER PPP & PLUS projects (H2020)**
    - WP leaders
    - Steering Committee members
  - after detailed review of existing work => starting from scratch

# eLTER Data Infrastructure – Services



# eLTER CI – selected services for users

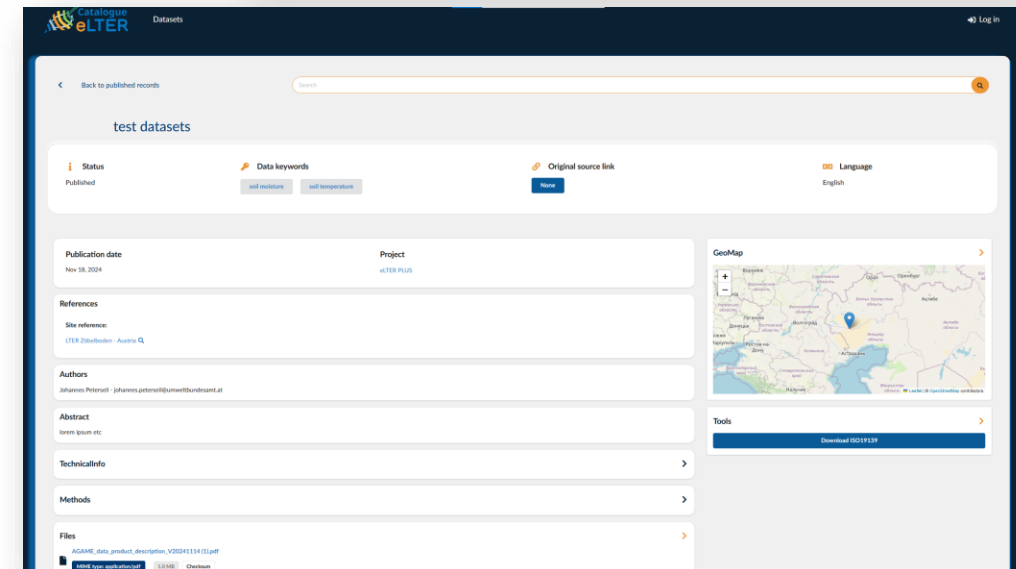
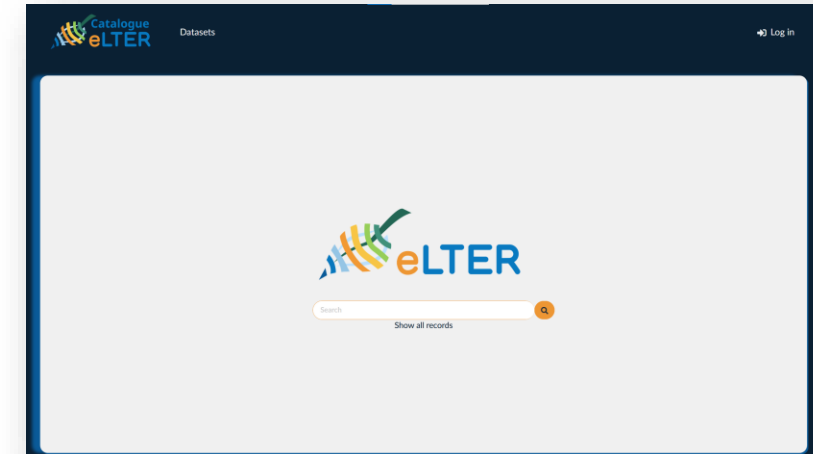
**Main eLTER services for users** (researchers, data analysts, common users):

- data registry & catalog services
  - ingested data & metadata
    - including harvesting of partner registries
  - **close and successful cooperation with EOSC CZ NRP**
- data view & analysis environment
  - DataLabs – cooperation with e-INFRA CZ (CERIT-SC)
  - spatial data analysis, time-series data analysis, data visualization portal, etc.
- supporting services
  - service portal („marketplace“)
  - vocabularies and ontologies
- etc.

# Registry & Catalog Services

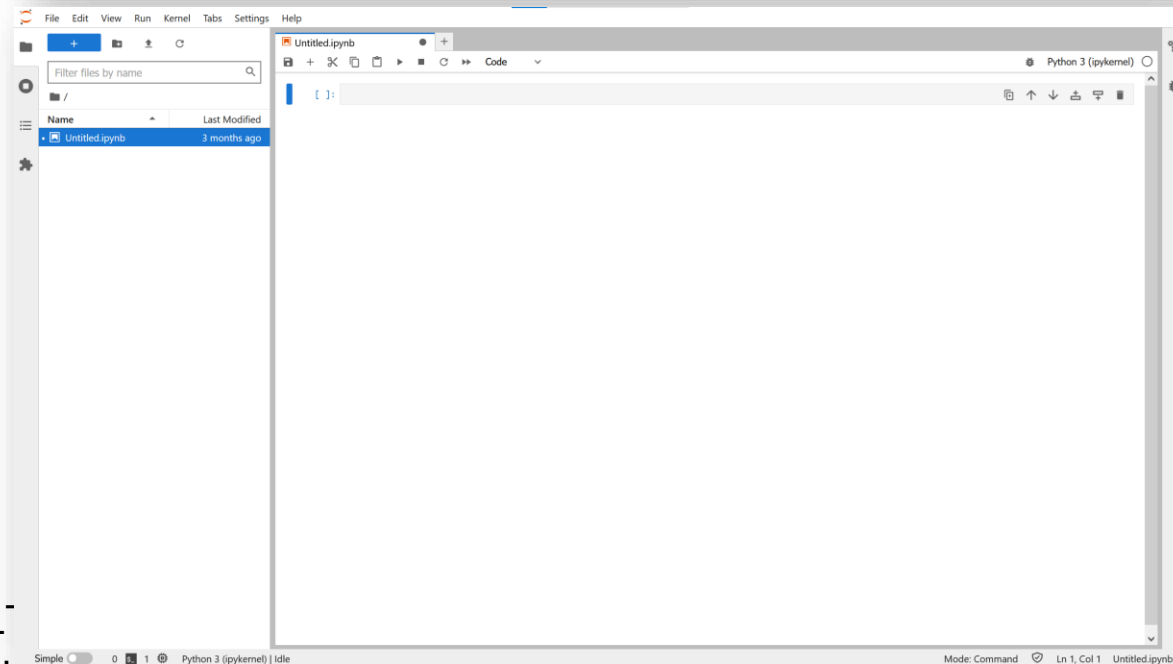
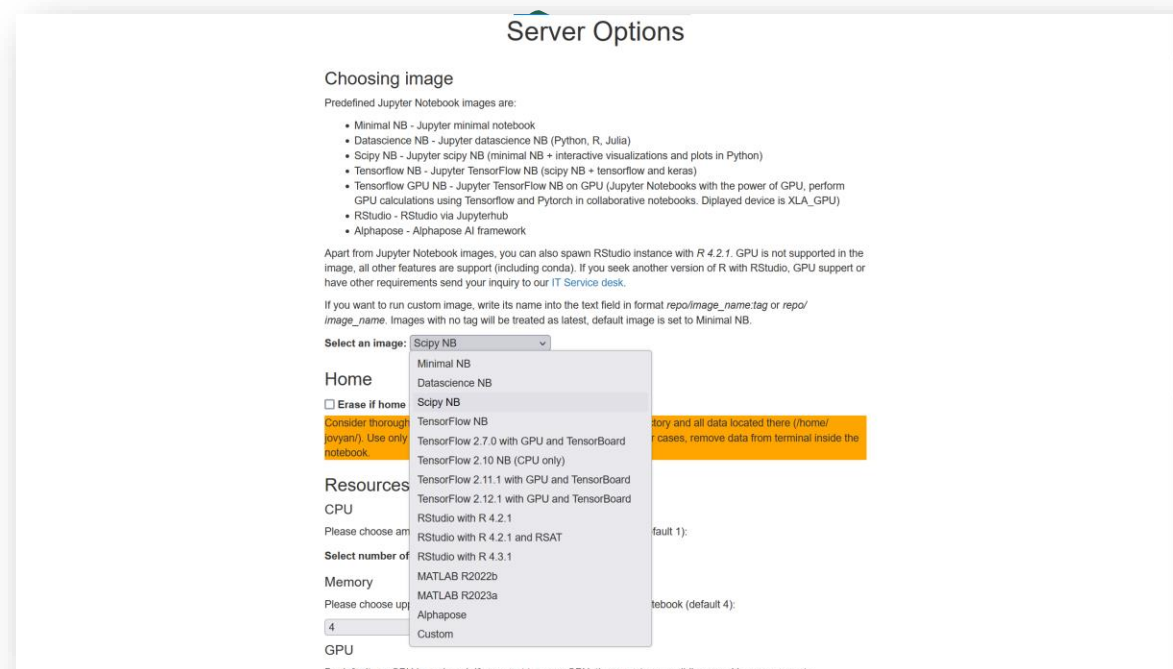
## Digital Asset Register (DAR)

- pilot version deployed, prepared for production
  - **employs the EOSC NRP solution**
    - close and successful collaboration
- comprises of two sub-components:
  - **Registry**
    - includes all the eLTER's data and descriptive metadata
      - but also other assets - FAIR
    - provides PIDs, rich metadata model, data backups etc.
    - includes the **data ingest functionality**
  - **Catalog**
    - a single point describing all the datasets available to the community
      - including external ones
    - harvests metadata from various dataset registry services
      - like B2SHARE, SITES, ZENODO, etc.



# DataLabs

- data processing/analysis environment
  - based on JupyterHub
  - via web browser
- expertly crafted images available
  - by CERIT-SC experts
- libraries for easier access to the data
  - Python, R, CLI scripts, etc.
  - will also enable to load external datasets





# DEIMS-SDR, eLTER DiV, Vocabs

## DEIMS-SDR

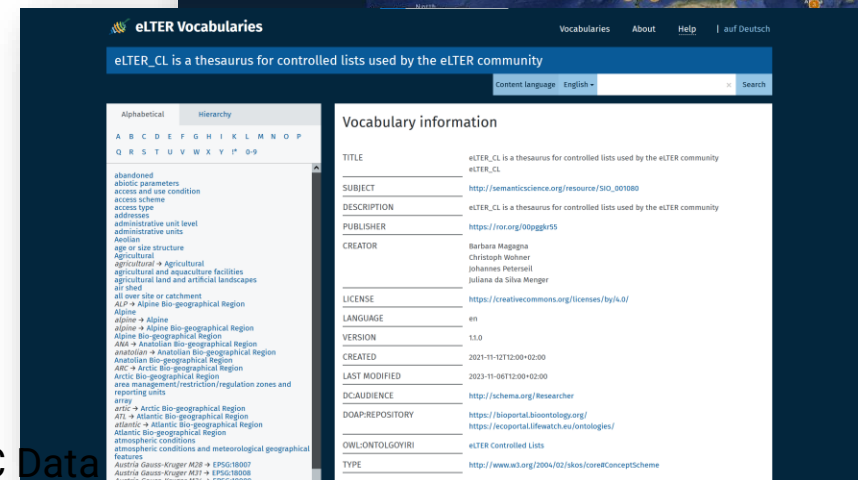
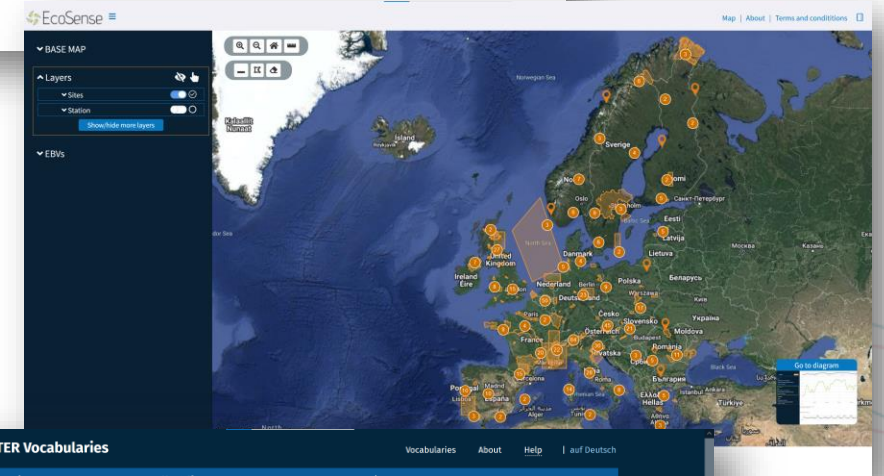
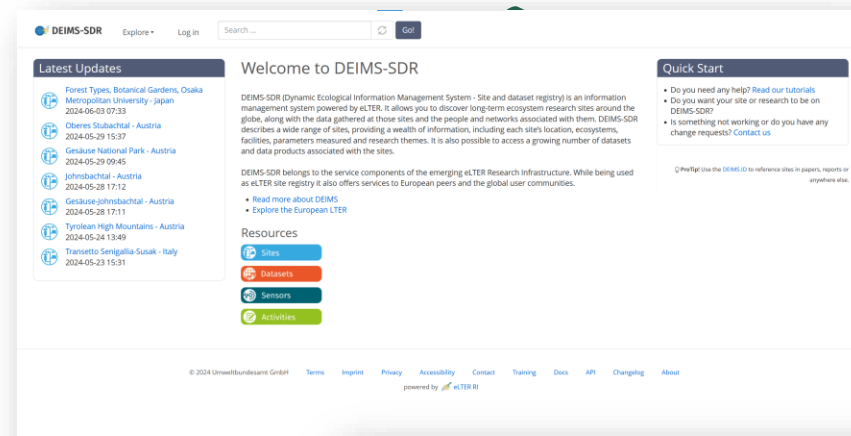
- information about eLTER sites

## eLTER DiV (*Data and Integrated Visualizations*)

- exploration of available datasets and their visualizations

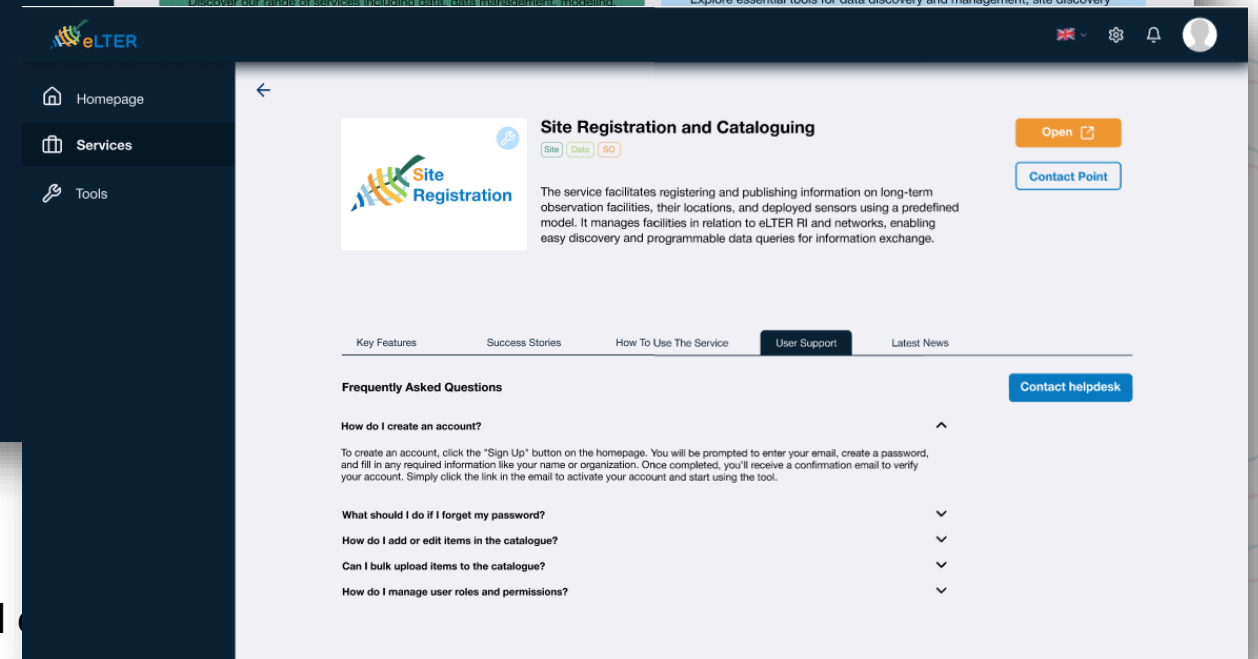
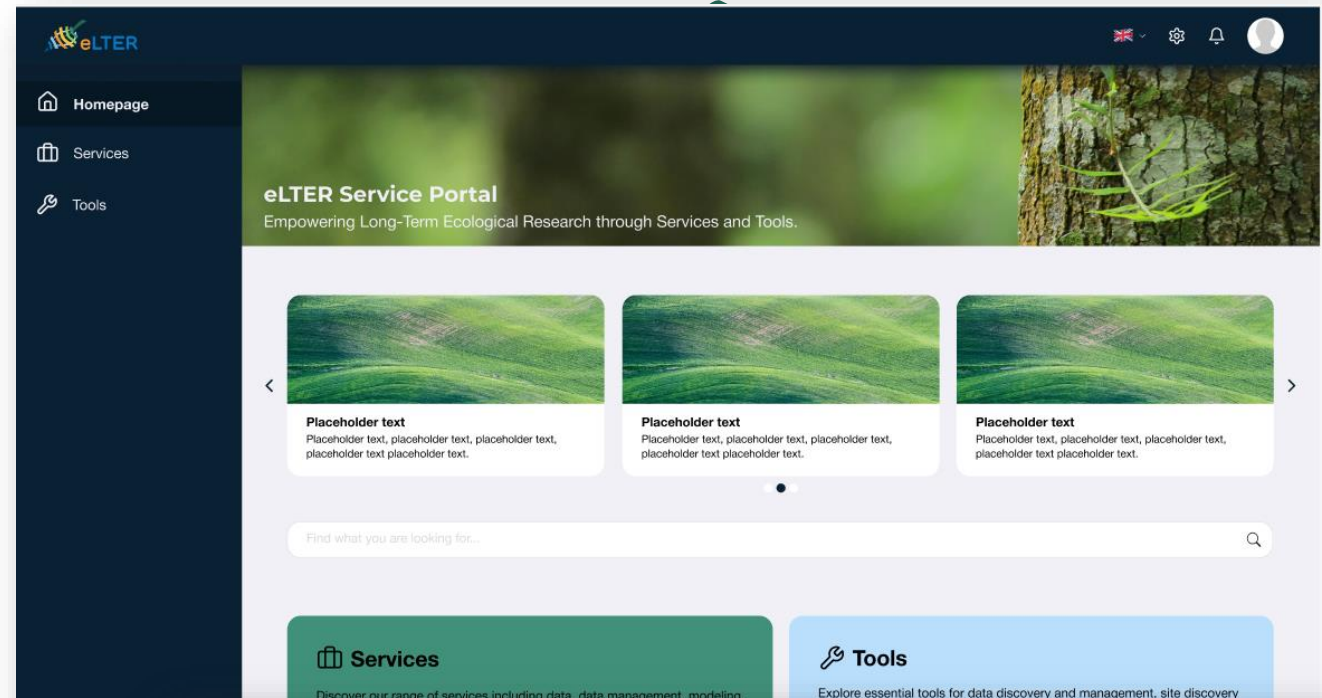
## eLTER Vocabularies

- eLTER controlled vocab lists based on SKOSMOS



# eLTER Service Portal

- a list of eLTER tools & services
  - information (per tool/service):
    - the service/tool leaflet
    - info about author
    - related (scientific) results
    - questions/answers
    - tutorials
  - queries, tags, categories, ...
- a community-supporting portal
  - discussion forums
    - (per tool/service or general)
  - users' ratings? feedback forms?



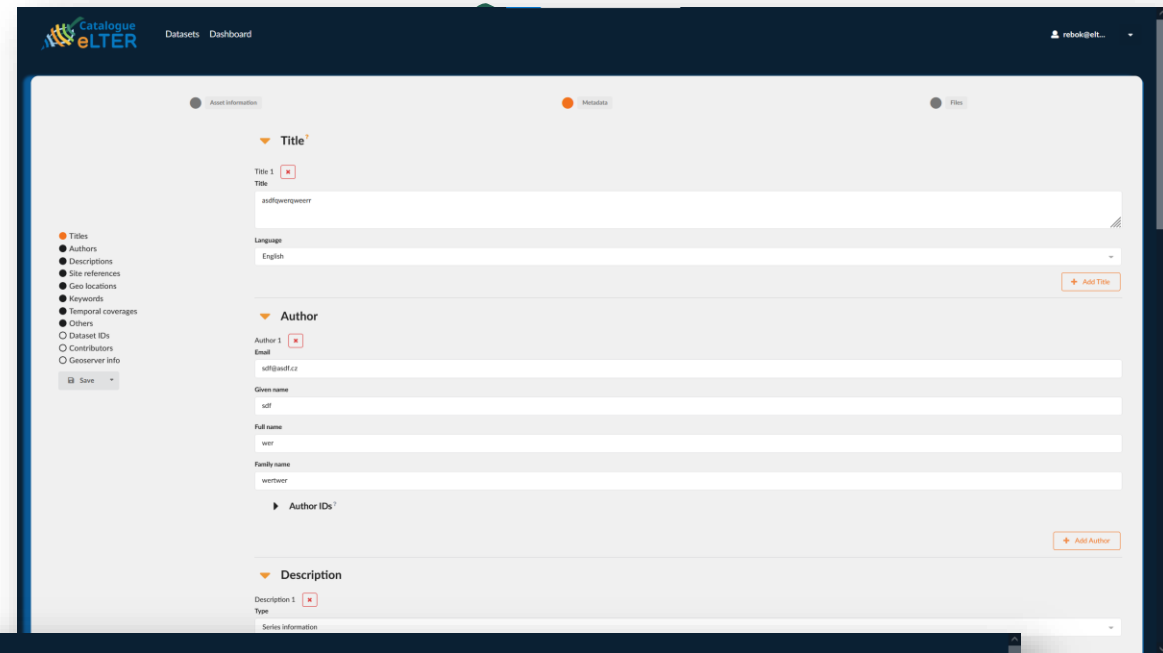
# eLTER CI – selected services for data providers

## Main eLTER services for data providers:

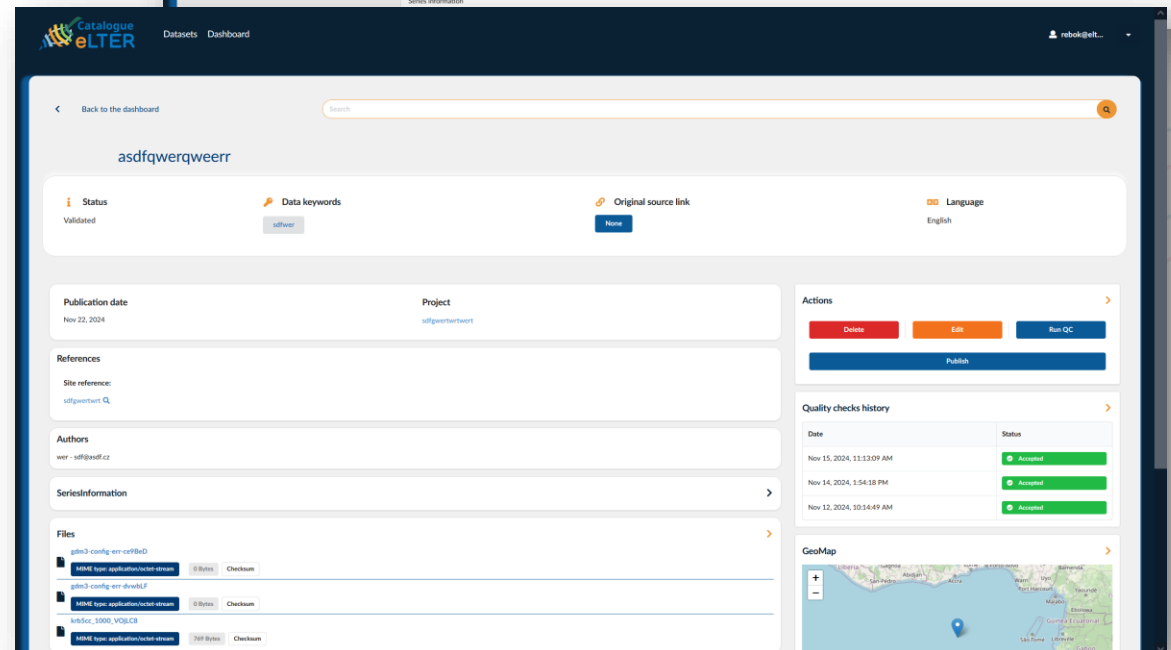
- data ingest portal
  - ability to ingest records (data + metadata) and control their publishing
- complex quality checking mechanisms
  - basic (metadata checks) as well as very complex quality checking procedures
    - including the on-site support (before providing the data)
- extended data provisioning services
  - automated data provisioning/loading by/from external data infrastructures
- monitoring the use of provided datasets
  - includes data provenance mechanisms

# Data Ingest Portal (@ eLTER DAR)

- a full control over ingested datasets
  - submitting metadata & data of records
    - publish & approve processes
  - overview of QA/QC testing results
    - including exploration of logs if declined
- in production from Jan'25



The screenshot shows the 'Catalogue eLTER' dataset creation interface. It features a sidebar with a navigation menu including: Titles, Authors, Descriptions, Site references, Geo locations, Keywords, Temporal coverages, Others, Dataset IDs, Contributors, and Geoserver info. The main form is divided into sections: 'Title' with a 'Title 1' field containing 'asdfqwerqweerr' and a 'Language' dropdown set to 'English'; 'Author' with an 'Author 1' field containing 'sd@asd.cz', 'Given name' 'sd', 'Full name' 'wer', and 'Family name' 'westhaer'; and 'Description' with a 'Description 1' field. There are 'Add Title', 'Add Author', and 'Save' buttons.



The screenshot shows the dataset overview page for 'asdfqwerqweerr'. It includes a search bar, a 'Back to the dashboard' link, and a 'Status' indicator (Validated). Key information includes 'Data keywords' (sdwer), 'Original source link' (New), and 'Language' (English). The page is divided into several sections: 'Publication date' (Nov 22, 2024), 'Project' (sdqwerqweerr), 'References' (Site reference: sdqwerqweerr), 'Authors' (wer - sd@asd.cz), 'Series information', and 'Files' (listing files with MIMe type and Checksum). On the right, there are 'Actions' (Delete, Edit, Run QC, Publish) and 'Quality checks history' (listing checks with dates and 'Accepted' status). A 'GeoMap' is also visible at the bottom right.

# eLTER Data Infrastructure – Summary



# eLTER Cyberinfrastructure (Data Infrastructure)

- a complex and powerful data infrastructure supporting European Research Community of Environmental Sciences
  - a Central Facility of (hopefully) future ERIC
    - Czech Republic will apply for hosting (ongoing discussions & preparations, including MŠMT)
- the goals:
  - to **support data providers** with a convenient environment for their data management, including quality assurance mechanisms and usage overview
  - to **support data users** with a convenient and powerful environment that enables various methods of complex data analysis and presentation of achieved results, over high-quality and easily findable/accessible datasets

# Opportunities for Czech environmental community

- **European infrastructure designed and developed in the Czech Republic**
  - prestige, increased awareness of our national activities
    - i.e. "we know how to build the infrastructures and analyse data"
  - our ambition is to be partners also in the field of advanced data analysis
- **New opportunities for environmental scientists**
  - who has the data and can exploit it, has the knowledge (and thus has the advantages)
  - our long-term ambition is to build a research center interconnecting environmental scientists with IT scientists
- **Building and expanding European partnerships**
  - following the example of the ICOS Carbon Portal, the partners managing the data services become partners in the projects of the community
  - partnership/cooperation opportunities with the ENVRI-FAIR community (12 large RIs)
- **and surely many others**
  - for example, ...

**National Consortium Agreement eLTER-CZ in the process of signing**

(9 Czech institutions)

T. Rebok - NRP and eLTER-ERIC Data

Infrastructure

# A solution for other RIs/institutes

## The eLTER data infrastructure builds on our generic data analysis solution

- powerful data-pipeline oriented approach
- strong automation via modern GitOps (deployment as well as management)

## The employed automation allows to run and manage multiple instances

- the CI can be deployed and managed on demand for other RIs/institutes
  - adapted to local needs (deployed locally)
  - currently in discussion with 2 universities
- can provide a local data management and analysis solution compatible with the centralized EOSC infrastructure and services
  - compatible metadata models, quality checks, tools deployed, ...

Automated CI deployment and management

