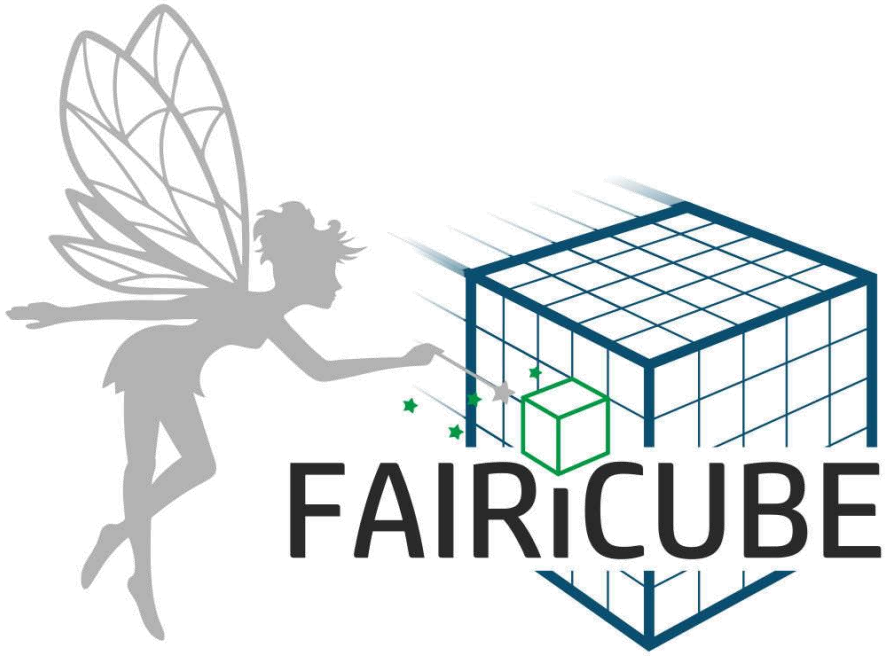


FAIRICUBE
F.A.I.R. INFORMATION CUBES

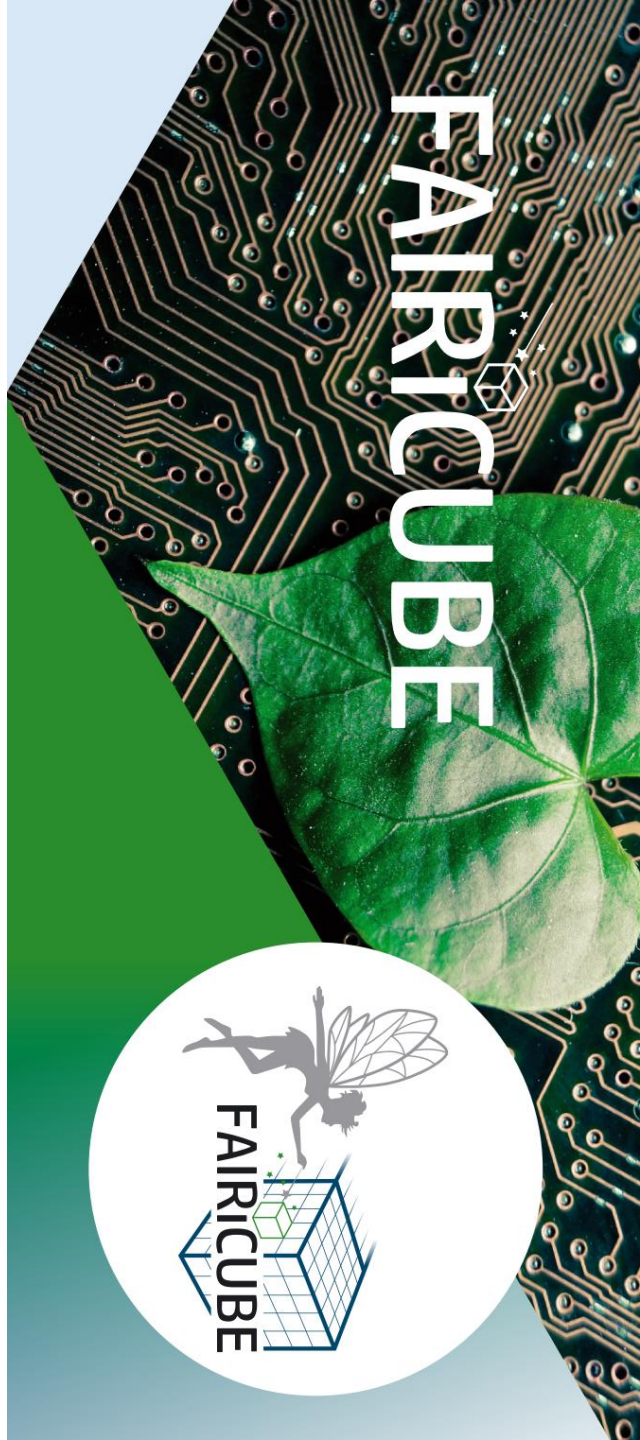
OUTLOOK ON THE GREEN DEAL DATA SPACE (GDDS)

KATHI SCHLEIDT & FAIRICUBE PARTNERS, 1.12.2025

FAIRICUBE INTRODUCTION

Stefan Jetschny, **Kathi Schleidt**

- FAIRiCUBE Introduction and Objectives
- Selected Use Cases
- FAIRiCUBE Hub services
- Validation framework
- AI Ethics
- Input for Green deal data spaces



OBJECTIVES

- enable players from beyond classic Earth Observation (EO) domains to provide, access, process, and share gridded data and algorithms in a FAIR and TRUSTable manner *(SO1)*
  - making Earth Observation (EO) data more accessible, interoperable, and impactful *(SO1)*
  - creating FAIRiCUBE HUB, a crosscutting platform and framework for data ingestion, provision, analysis, processing, and dissemination *(SO1, SO2, SO3)*
    - Building on existing technology and services
- Exploit and leverage machine learning techniques *(SO2)*
  - Pilot studies to integrate ML capabilities into FAIRiCUBE Hub services (Lab)
- Demonstrate and validate FAIRiCUBE services as iterative co-creation process with Use Cases *(SO3, SO4)*
  - Real world problems in close collaboration with stakeholders
- Provide persistent input for the design and implementation of the green deal data space (SO5)

CONSORTIUM

- Research institutes
  - NIL - NILU climate and environmental research institute, Norway
  - WER - Wageningen university and research, Netherlands
  - NHM - Natural History Museum Vienna, Austria
- Environmental SME's
  - S4E - space4environment, Luxembourg
  - 4SF - 4sfera, Spain
  - EPS - Epsilon, Italia
- Infrastructure service providers
  - EOX - EOX, Austria
  - CUB - Constructor University Bremen, Germany, supported by rasdaman GmbH

domain specialist/use case owner, geospatial data specialist, infrastructure specialists

THIS IS FAIRICUBE

2022

60 on Teams
40 on GitHub
270 on LinkedIn
400 on Website

2025

We brought together technical specialists and researchers from various domains.
We start understanding each other and contribute to FAIRiCUBE Hub!

FAIRICUBE USE CASES

UC1 : Urban adaptation to climate change (urban focus)
Space4Environment (S4E), Luxembourg

UC2: Agriculture and biodiversity nexus (regional focus)
Wageningen Environmental Research (WER), Niederlande

UC3: Environmental adaptation genomics in drosophila (regional focus)
Naturhistorisches Museum (NHM), Wien, Österreich

UC4: Spatial and temporal assessment of neighborhood building stock (urban focus)
NILU (Norwegian Institute for Air Research), Norwegen

UC5: Biodiversity occurrence Cubes (regional to European focus)
Naturhistorisches Museum (NHM), Wien, Österreich

FAIRICUBE USE CASE 1

„URBAN ADAPTATION TO CLIMATE CHANGE"

Maria Ricci, Manuel Löhnertz, Marco Cattaneo, Andrea Peters
Space4Environment, Luxembourg

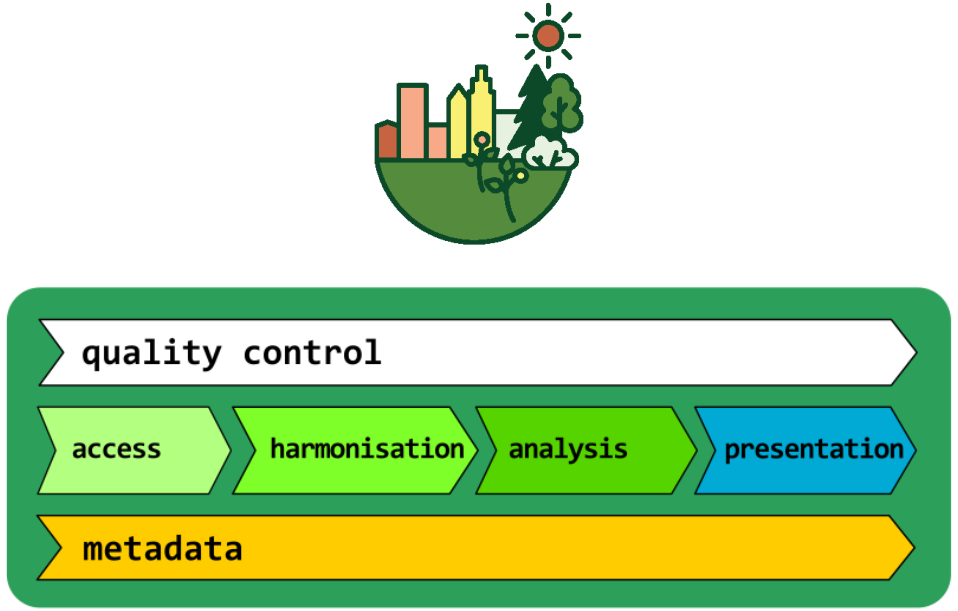
EUROPEAN AND LOCAL RESEARCH QUESTIONS

European Level Strategy

- Can we find common patterns in European cities' climate adaptation strategies using datasets like Copernicus and Eurostat?
- How do measures such as increasing urban tree cover affect climate adaptation across European cities?

Local Focus on Luxembourg City

- What is the correlation between invasive plant species and environmental factors in Luxembourg City?
- Findings aim to support mitigation actions by prioritizing areas at higher risk



European cities

Luxembourg city

DATA HARMONIZATION AND ANALYSIS

Data Harmonization

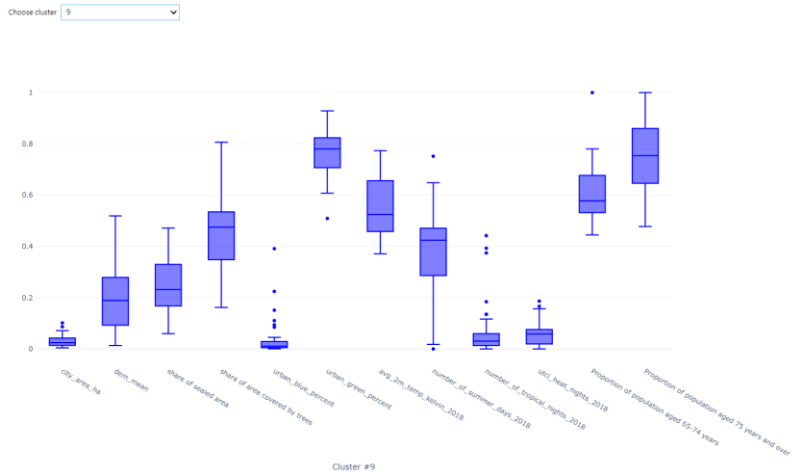
- We combined heterogeneous datasets by aligning projections, locations, and timeframes into structured spatial data cubes.

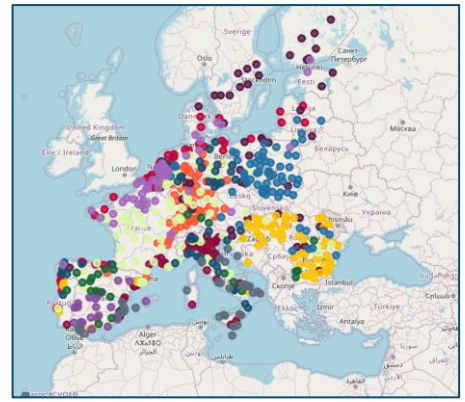
Analysis

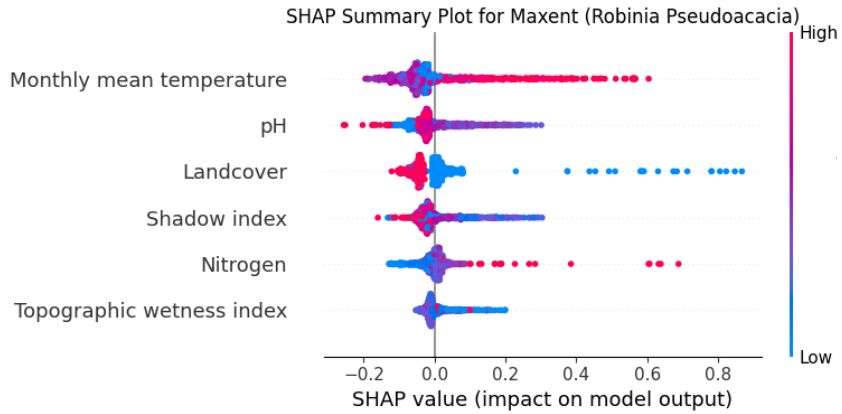
- We tested BLSTM and Gradient Boosting to estimate missing data from correlated features.
- We grouped cities by land climate and socioeconomic characteristics using clustering algorithms.

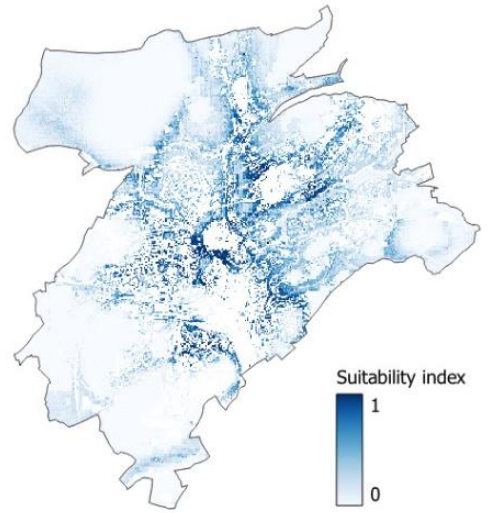
Species Distribution & Impact Assessment

- We used the MaxEnt model to predict invasive species habitats and SHAP to explain environmental impacts.









BLSTM: Bidirectional Long Short-Term Memory; Maxent: Maximum Entropy model; SHAP: SHapley Additive exPlanations

SUPPORT FROM FAIRICUBE INFRASTRUCTURE

Data Preprocessing and workflow sharing

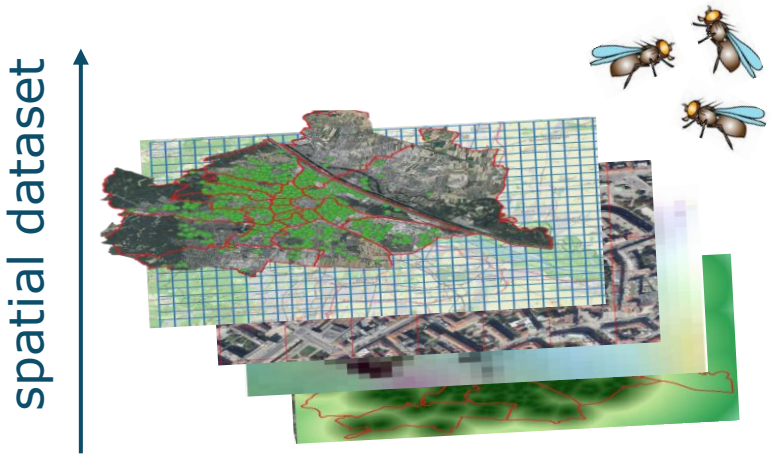
- EOX Lab enabled data tasks like reprojecting datasets, rasterizing vector data, and deriving environmental indices.
- EOX Lab hosted notebooks facilitating reproducible research and workflow sharing across the team.

Fast prototyping

- The modular infrastructure allowed to quickly test different models and software solutions

Use case synergies

- We benefited from the diverse domain expertise and often contributed to each other use cases



Urban cube for Vienna (UC3)

BENEFITS AND DRAWBACKS OF FAIRICUBE SERVICES

Interactive Communication Tools

- Grafana, Streamlit apps and FAIRiCUBE scrollytelling maps help dissemination of project results.

Centralized Datasets via STAC Catalog

- The FAIRiCUBE STAC Catalog and Zenodo keep track of data and analysis/processing metadata.
- Metadata editing is time-consuming

Validation

- Validation conceptual framework helped early discovery of data issues
- Systematic validation slows down product development

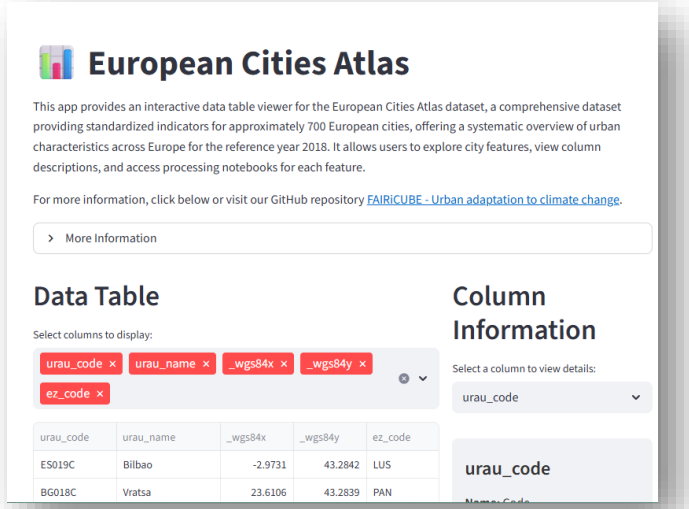
ACHIEVEMENTS AND IMPACT

European cities atlas

- City features data cube combining climate, land use, and socio-economic indicators across Europe.
- Openly available and easy to explore
- Interactive Notebooks to reproduce cluster analysis

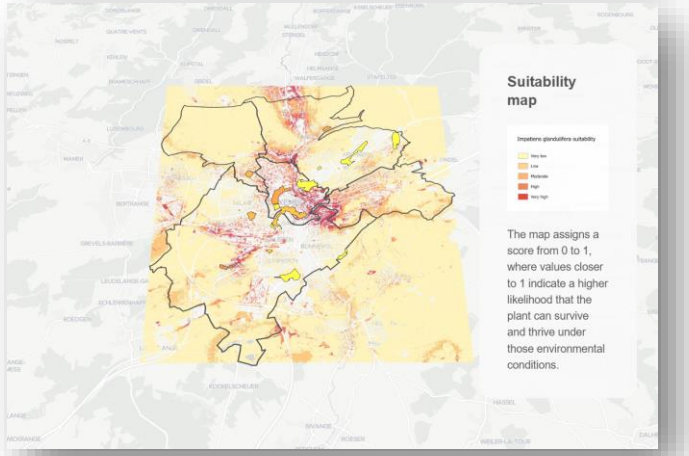
Local Urban Monitoring

- High-resolution habitat cube in Luxembourg City monitored invasive plant species and informed ecological risk management.
- Explainable AI tool SHAP revealed environmental drivers affecting species distribution for better decision making.



Indicators explorer

euopean-cities-atlas.streamlit.app



Scrollytell map

uc1.fairicube.nilu.no

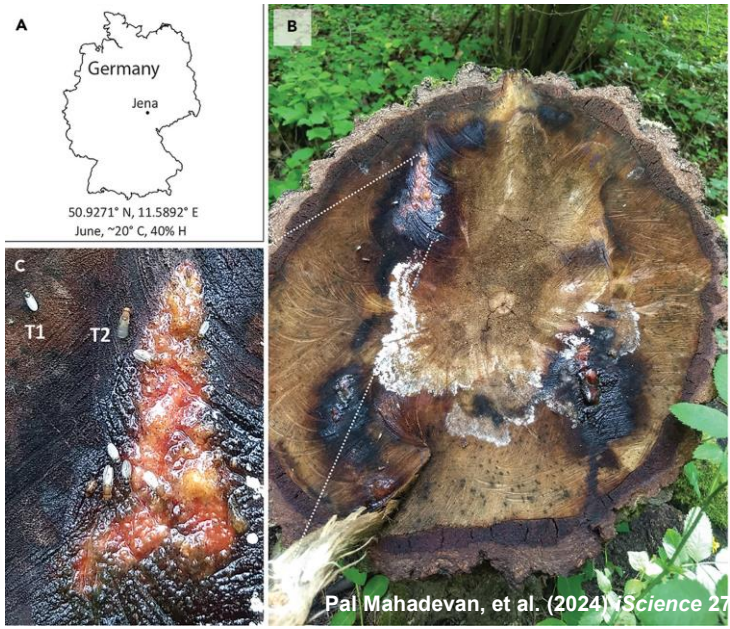
FAIRICUBE USE CASE 3

„ENVIRONMENTAL ADAPTATION GENOMICS IN DROSOPHILA"

Martin Kapun, Sonja Steindl
NHM Wien

DROSOPHILA AS AN ECO-EVO MODEL

- Approximately 1,600 species
- From local food specialists to globally distributed generalists
- **Ecology and environmental selective forces that determine adaptation largely unknown**
- **Genes underlying environmental adaptation often unknown**



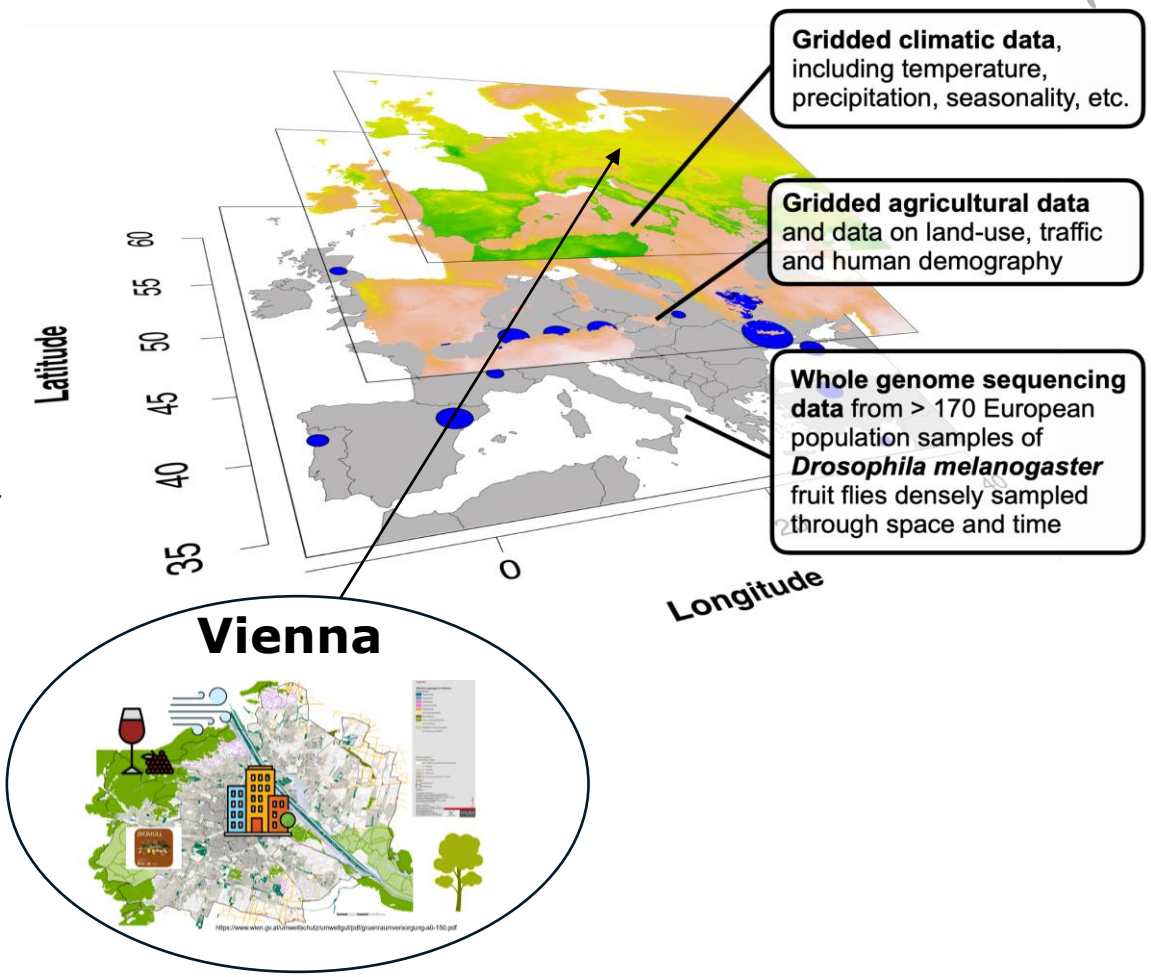
RESEARCH QUESTIONS

1. Continental scale

- Which <u>environmental factors have the strongest influence</u> on adaptation in *D. melanogaster.*
- Which <u>genomic regions or genes in *D. melanogaster*</u> show signatures of adaptation to specific environmental conditions?

2. Local (urban) scale

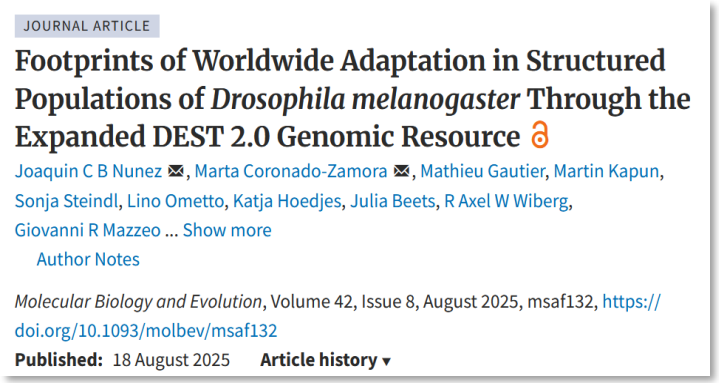
- How does the <u>species richness and composition change</u> in urban and rural habitats?
- Which <u>environmental factors influence</u> the occurrence of certain species?



1. ADAPTIVE EVOLUTION IN EUROPE

Methods

- **Genomic Datasets**
  - Pan-European fly collections & genome sequencing
  - >300 Populations-samples through time and space
  - Published in MBE (IF: 10.7)
- **Imputation of missing genomic datasets**
  - ML methods; Collaboration with UC4
  - Paper in IEEE Xplore; 2nd paper just accepted
- **Environmental datasets**
  - Querycube online tool to slice E/O data from datacubes stored at Rasdaman
  - Climate Data (ERA5) through Copernicus API
  - Interpolated climate data from WorldClim II



JOURNAL ARTICLE

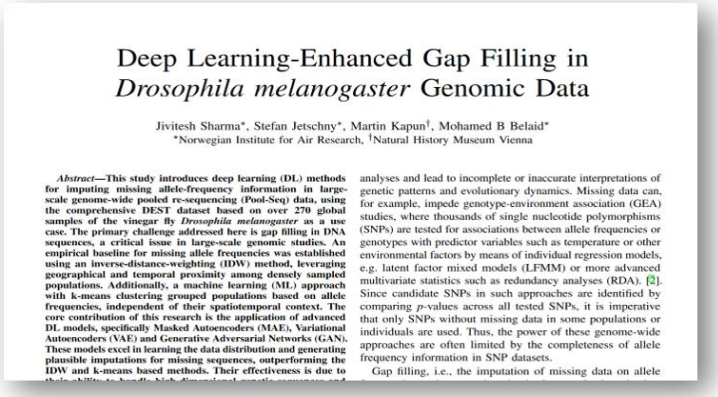
Footprints of Worldwide Adaptation in Structured Populations of *Drosophila melanogaster* Through the Expanded DEST 2.0 Genomic Resource

Joaquin C B Nunez, Marta Coronado-Zamora, Mathieu Gautier, Martin Kapun, Sonja Steindl, Lino Ometto, Katja Hoedjes, Julia Beets, R Axel W Wiberg, Giovanni R Mazzeo ... Show more

Author Notes

Molecular Biology and Evolution, Volume 42, Issue 8, August 2025, msaf132, https://doi.org/10.1093/molbev/msaf132

Published: 18 August 2025 **Article history**



Deep Learning-Enhanced Gap Filling in *Drosophila melanogaster* Genomic Data

Jivitesh Sharma*, Stefan Jetschny*, Martin Kapun†, Mohamed B Belaid*
*Norwegian Institute for Air Research, †Natural History Museum Vienna

Abstract—This study introduces deep learning (DL) methods for imputing missing allele-frequency information in large-scale genome-wide pooled re-sequencing (Pool-Seq) data, using the comprehensive DEST dataset based on over 270 global samples of the vinegar fly *Drosophila melanogaster* as a use case. The primary challenge addressed here is gap filling in DNA sequences, a critical issue in large-scale genomic studies. An empirical baseline for missing allele frequencies was established using an inverse-distance-weighting (IDW) method, leveraging geographical and temporal proximity among densely sampled populations. Additionally, a machine learning (ML) approach with k-means clustering grouped populations based on allele frequencies, independent of their spatiotemporal context. The core contribution of this research is the application of advanced DL models, specifically Masked Autoencoders (MAE), Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN). These models excel in learning the data distribution and generating plausible imputations for missing sequences, outperforming the IDW and k-means based methods. Their effectiveness is due to

analyses and lead to incomplete or inaccurate interpretations of genetic patterns and evolutionary dynamics. Missing data can, for example, impede genotype-environment association (GEA) studies, where thousands of single nucleotide polymorphisms (SNPs) are tested for associations between allele frequencies or genotypes with predictor variables such as temperature or other environmental factors by means of individual regression models, e.g. latent factor mixed models (LFMM) or more advanced multivariate statistics such as redundancy analyses (RDA). [2]. Since candidate SNPs in such approaches are identified by comparing p-values across all tested SNPs, it is imperative that only SNPs without missing data in some populations or individuals are used. Thus, the power of these genome-wide approaches are often limited by the completeness of allele frequency information in SNP datasets.

Gap filling, i.e., the imputation of missing data on allele

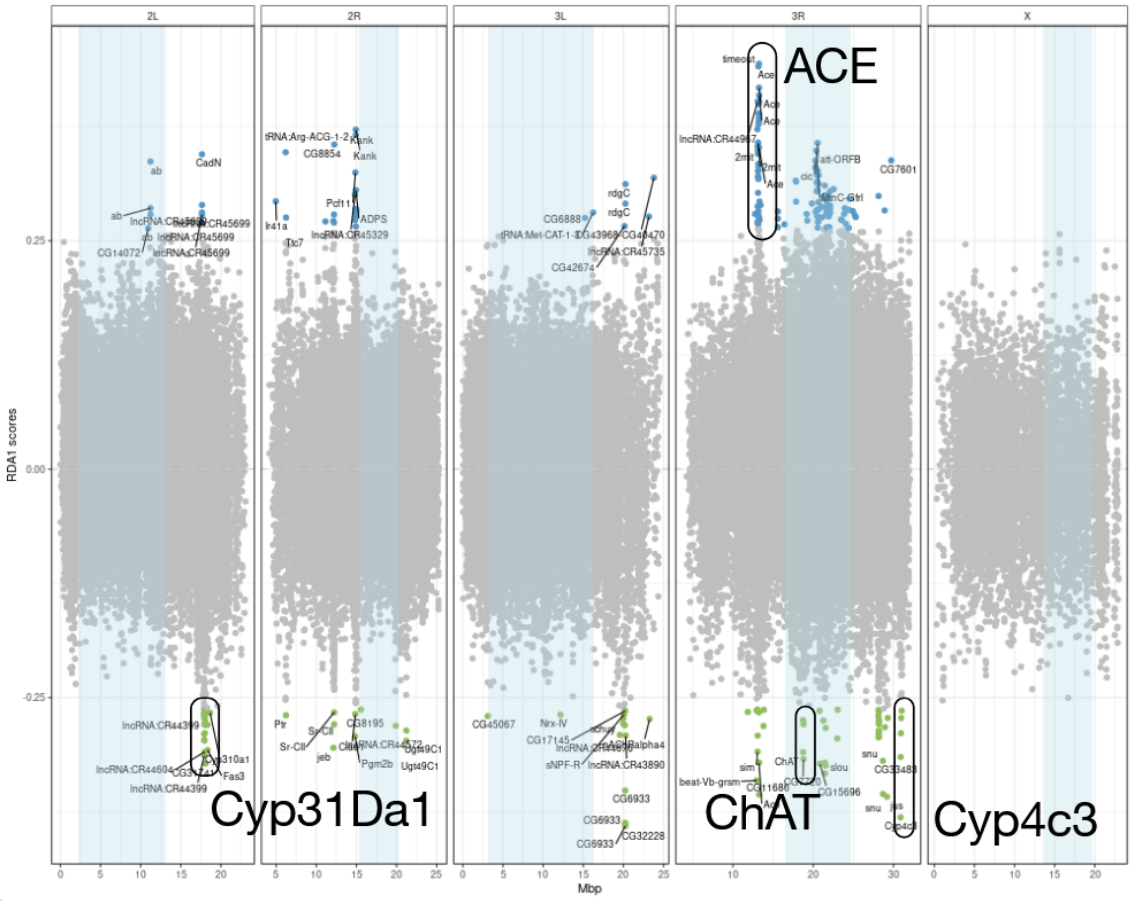
1. ADAPTIVE EVOLUTION IN EUROPE

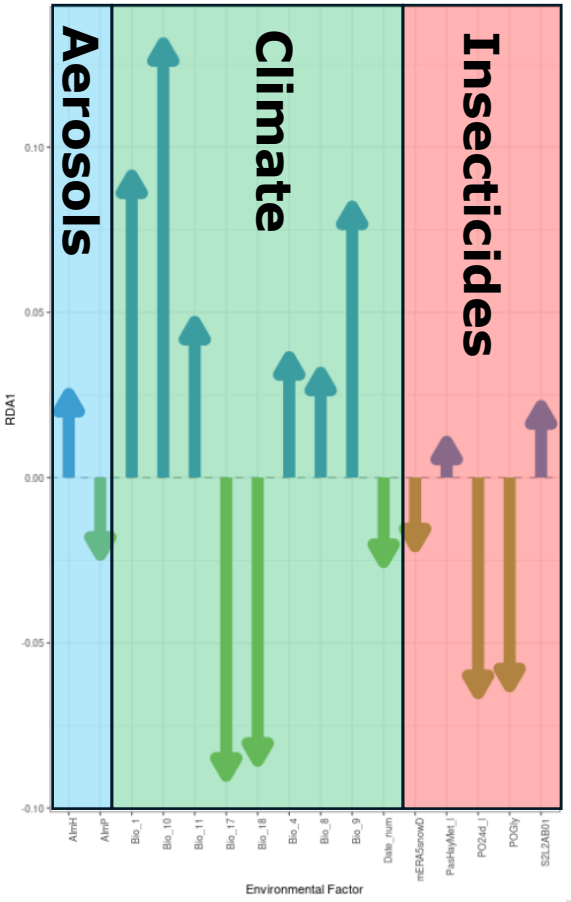
Linking genomic and E/O data

- 156 environmental variables
- 293 population samples
- 500,000 genome-wide polymorphisms (SNPs)
- Redundancy analysis (RDA) to test for associations

Results

- Many candidate genes targets of insecticides or with effect on stress resistance

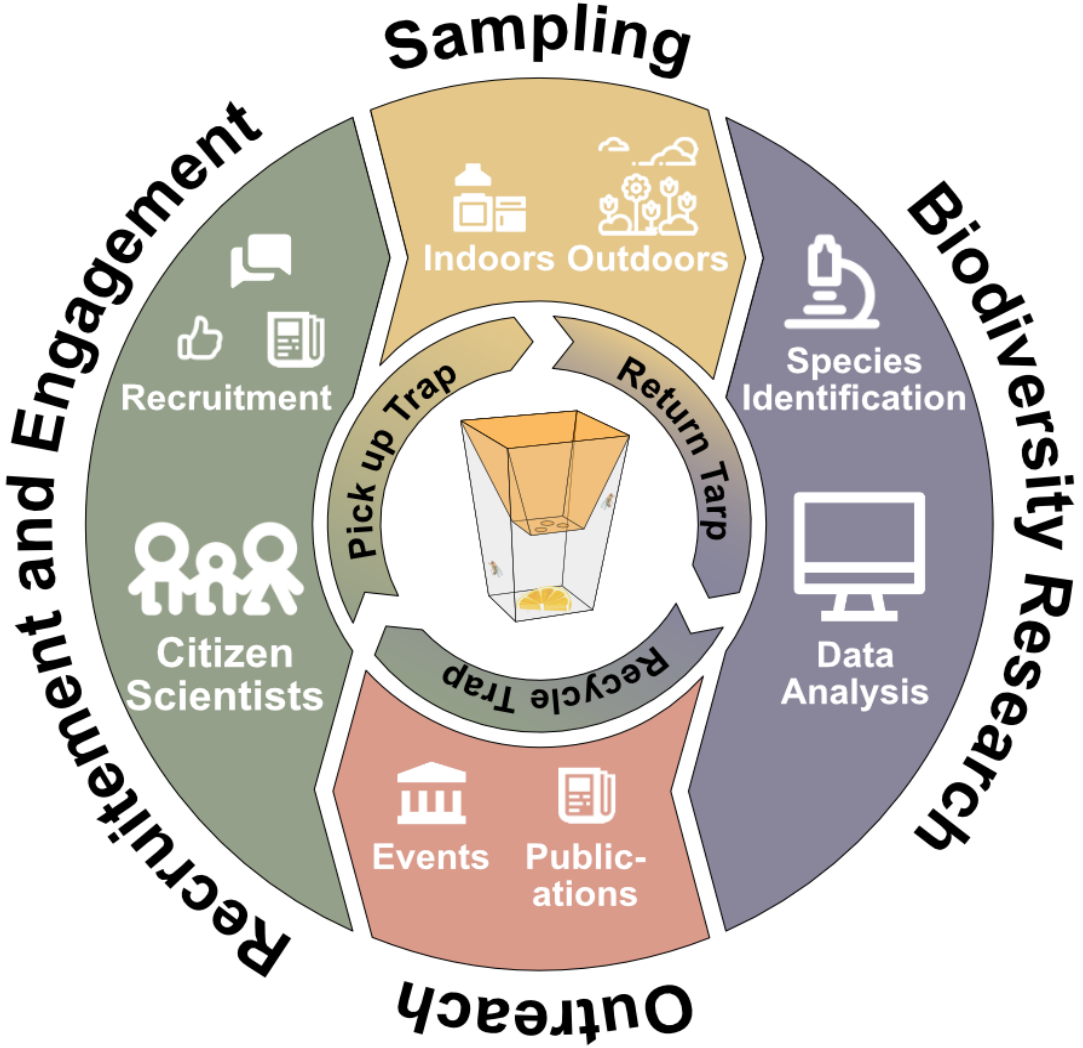




2. ECOLOGY OF URBAN *DROSOPHILA*

Methods

- Citizen Science
  - Data acquisition & education
- Environmental Data (n=58)
  - Climate Data from Geosphere Austria
  - Land-use data in collaboration with UC1
- Data Analysis
  - Redundancy Analysis
  - Species distribution modelling
- Genomic comparison to European flies

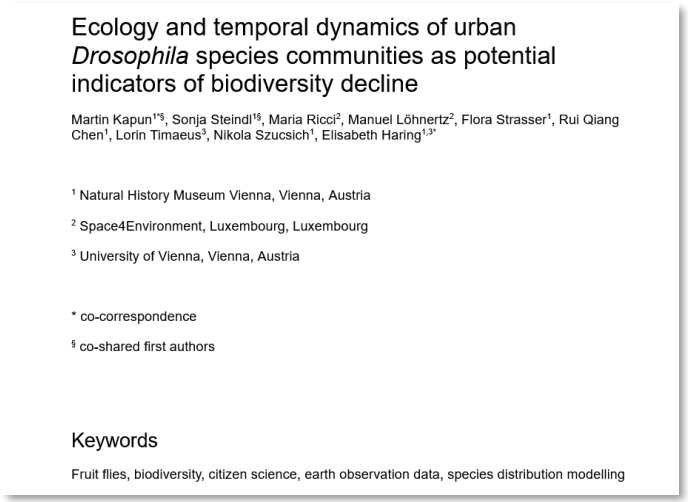


2. ECOLOGY OF URBAN *DROSOPHILA*

Results

- Ecology and Biodiversity
  - 168 citizen scientists; 278 traps; 18,231 flies
  - 13 species identified; 2 new for Austria
  - Characterization of ecological niches
  - 1 manuscript under review in Ecology and Evolution (IF:2.3)
- Public Education
  - Raising biodiversity awareness
  - Assessing motivation for participation
  - 1 manuscript under review in Ecology and Society (IF: 5.3)

https://doi.org/10.1101/2025.09.04.674250



Ecology and temporal dynamics of urban *Drosophila* species communities as potential indicators of biodiversity decline

Martin Kapun[1*§], Sonja Steindl[1§], Maria Ricci[2], Manuel Löhnertz[2], Flora Strasser[1], Rui Qiang Chen[1], Lorin Timaeus[3], Nikola Szucsich[1], Elisabeth Haring[1,3*]

[1] Natural History Museum Vienna, Vienna, Austria

[2] Space4Environment, Luxembourg, Luxembourg

[3] University of Vienna, Vienna, Austria

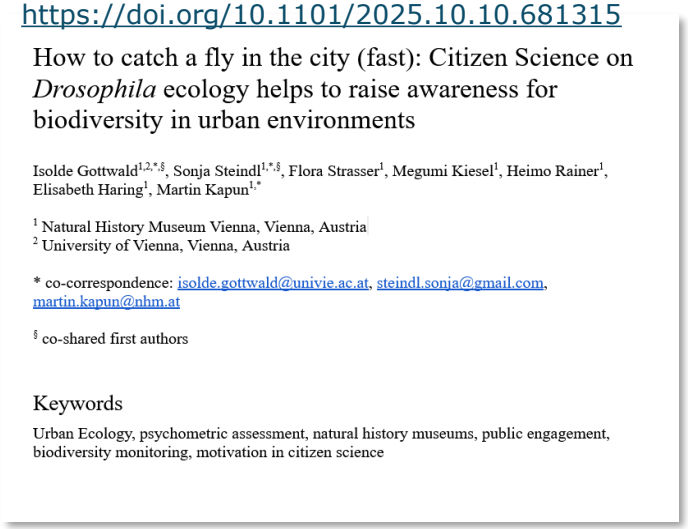
* co-correspondence

§ co-shared first authors

Keywords

Fruit flies, biodiversity, citizen science, earth observation data, species distribution modelling

https://doi.org/10.1101/2025.10.10.681315



How to catch a fly in the city (fast): Citizen Science on *Drosophila* ecology helps to raise awareness for biodiversity in urban environments

Isolde Gottwald[1,2,*,§], Sonja Steindl[1,*,§], Flora Strasser[1], Megumi Kiesel[1], Heimo Rainer[1], Elisabeth Haring[1], Martin Kapun[1,*]

[1] Natural History Museum Vienna, Vienna, Austria
[2] University of Vienna, Vienna, Austria

* co-correspondence: isolde.gottwald@univie.ac.at, steindl.sonja@gmail.com, martin.kapun@nhm.at

§ co-shared first authors

Keywords

Urban Ecology, psychometric assessment, natural history museums, public engagement, biodiversity monitoring, motivation in citizen science

SUPPORT FROM FAIRICUBE INFRASTRUCTURE

Data Preprocessing and workflow sharing

- EOX Lab enabled intersecting datasets from different domains and to test workflows.
- EOX Lab hosted Jupyter notebooks that facilitate documentation and sharing code and knowhow with partners.
- The digital library and the GitHub repositories facilitate reproducible research

Data access

- The ingested datasets in Rasdaman allowed access to previously unavailable datasets through our novel Querycube tool

Use case synergies

- Our research and work benefited from the specific domain expertise among the other FAIRiCUBE partners, which resulted in previously unprecedented and highly interdisciplinary ongoing research collaborations

BENEFITS AND DRAWBACKS OF FAIRICUBE SERVICES

Communication Tools

- FAIRiCUBE scrollytelling maps help dissemination of project results.
- Broad audience and media coverage through centralized social media campaigns
  - Providing content can by time-consuming

Data access & Documentation

- Support with access to domain knowledge and datasets that require specific knowledge for quality assessment
- Metadata catalog allows for comprehensive understanding and usage of datasets
  - Sometimes difficult to access metadata information

Validation

- Data and workflow validation as a key prerequisite to work with big datasets
  - Sometimes difficult to implement and uncertain standards

ACHIEVEMENTS AND IMPACT

Querycube

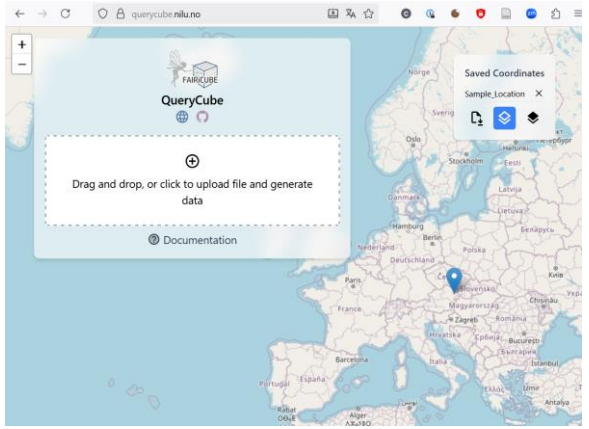
- An online tool to obtain point data for input coordinates from E/O datacubes stored at Rasdaman

Publications

- 3 accepted/published research papers, 2 papers under review; 1 in preparation
- Comprehensive E/O data allow unprecedented insights into environmental effects shaping genetic variation
- First time ecological inference of urban *Drosophila*
- New ML tools to impute incomplete genetic data

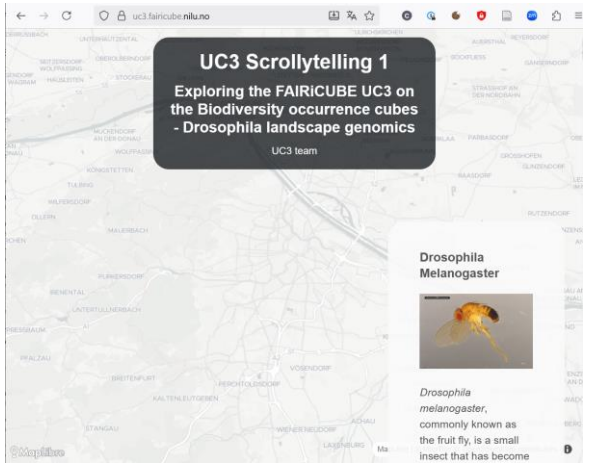
Outreach and Media

- Press articles, TV documentary, Museum workshops, Social medial posts



QueryCube tool

https://querycube.nilu.no/



Scrollytelling map

uc3.fairicube.nilu.no

FAIRICUBE HUB SERVICES
ADVANCES AND ENHANCEMENTS

Christian Schiller, Dimitar Mišev

- Evolution of services
  - FAIRiCUBE Lab (EOX/AWS)
  - Data Catalog and Browser
    - Evolving the STAC Standard and FAIRiCUBE Catalog Editor
  - rasdaman Workspace
  - rasdaman Dashboard
- Commercial outlook
  - Extended operation
  - Customers and supported Initiatives

FAIRiCUBE HUB

Services & Apps

Knowledge Base Services
- GPT4-0 ChromaDB Digital Library Query Tool
- Django, PostgreSQL
- GitHub, ReadTheDocs

rasdaman & OGC
- WM&, WMTS
- WCS, WCPS
- OAPI Cov.
- openEO

ML/AI
- MLflow
- TensorBoard
- DVC
- rasdaman AI model mgr

EDC GeoDB
- PostgREST API
- openEO (TBD)

SentinelHub
- Process API
- Batch API
- Stats API
- openEO
- WMS

etc.
- xarray
- s3
- etc.

Catalog
- GitHub Pull Requests
- GitHub Actions (CI/CD)
- GitHub Pages
- STAC-FastAPI
- STAC Browser

Data/Resource/ Code requests

FAIRiCUBE Lab - EOxHub Workspace

Worker Plane
- JupyterLab IDE
  - User Code Execution
- pygeoapi - Headless Execution
- GitHub code synchronisation
- Prometheus
- Grafana-Loki
- Docker Container

Control Plane
- Configuration Mgmt.
- GitHub Identity Provider
- JupyterLab Profiles
- Keycloak
- Shared Jupyter Notebooks
- Shared Conda Environments
- Shared Secrets
- Shared Object Storage
- Workload Orchestration
- Apps

Object Storage
- FAIRiCUBEs
- Models

FAIRiCUBE - rasdaman Workspace

- **Federated AI-Cubes**
- **Visual & zero-coding:** ChatCUBE, dashboard, openEO, WCS client, …
- **Coding:** Jupyter R+Python, OWSlib, …
- **Intelligent automatic import**
- **Pixel-level security**

Multi-parallel, distributed rasdaman server

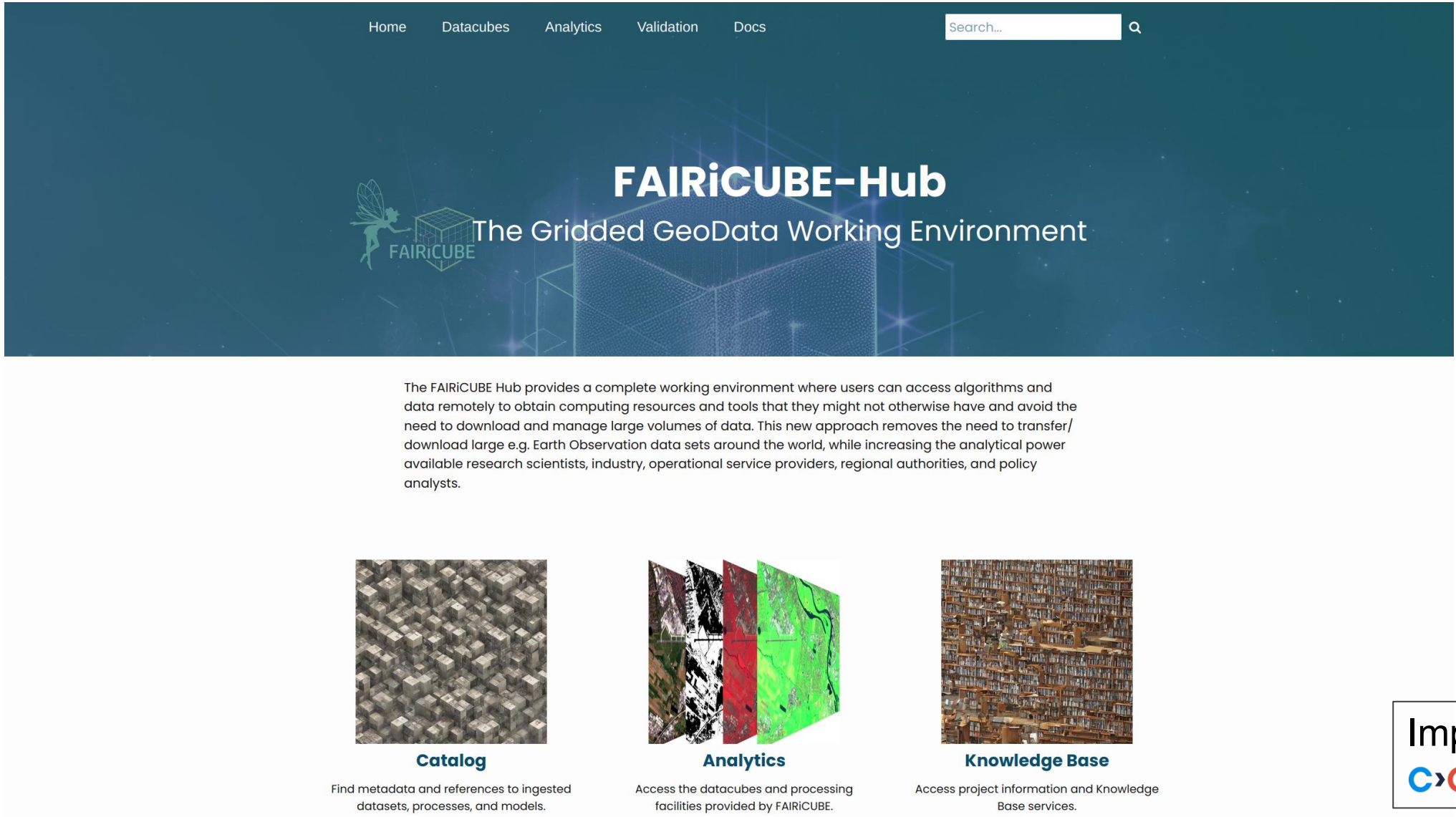
External archive integration
- GeoTIFF, NetCDF, …

rasdaman database
- FAIRiCUBEs
- Models
- 3rd Party Code (UDF)

Externals, e.g. Euro Data Cube and EarthServer Datacube Federation

OpenniCUBE

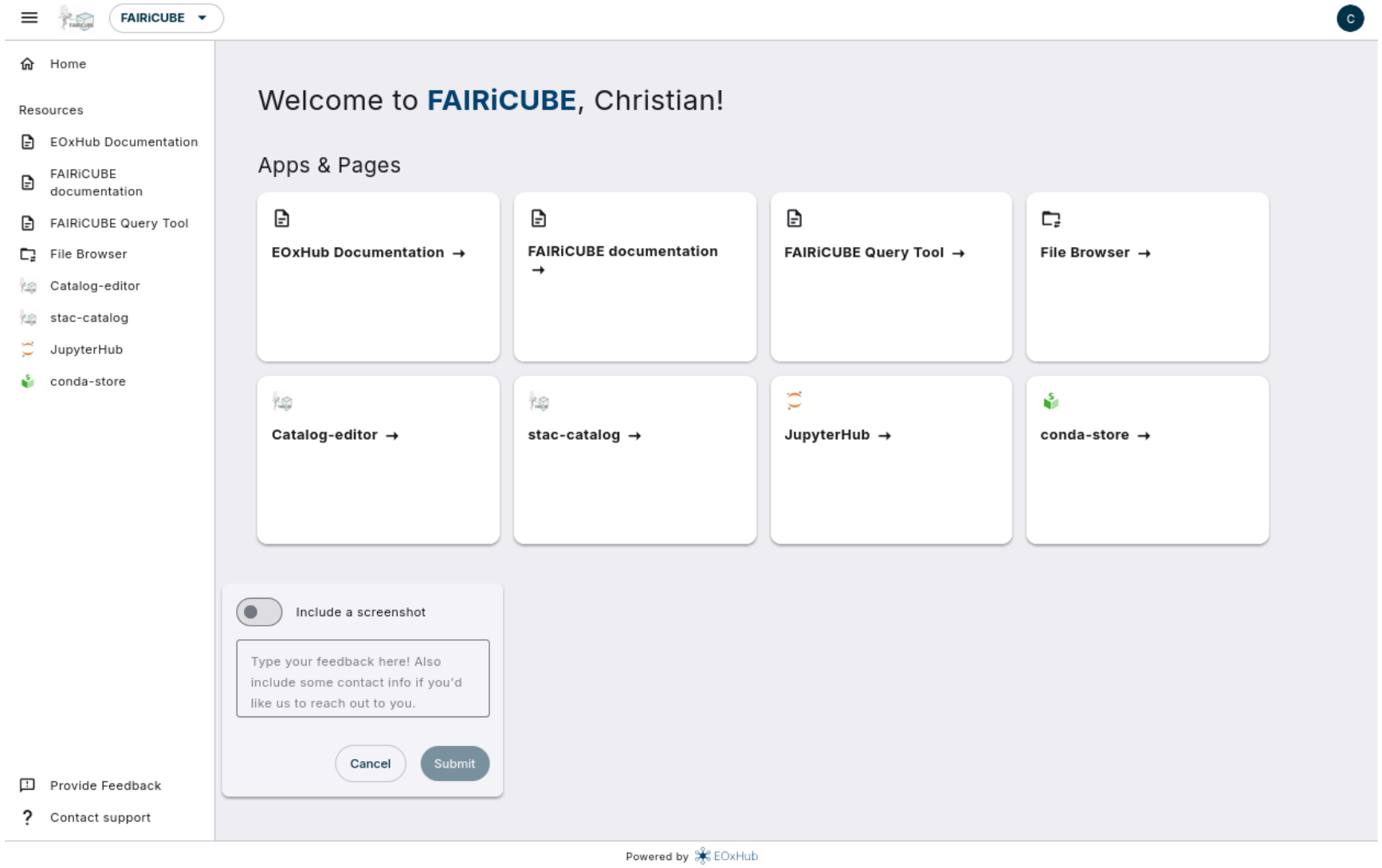
HUB.FAIRICUBE.EU



Implemented by CONSTRUCTOR

FAIRICUBE LAB – WORKSPACE

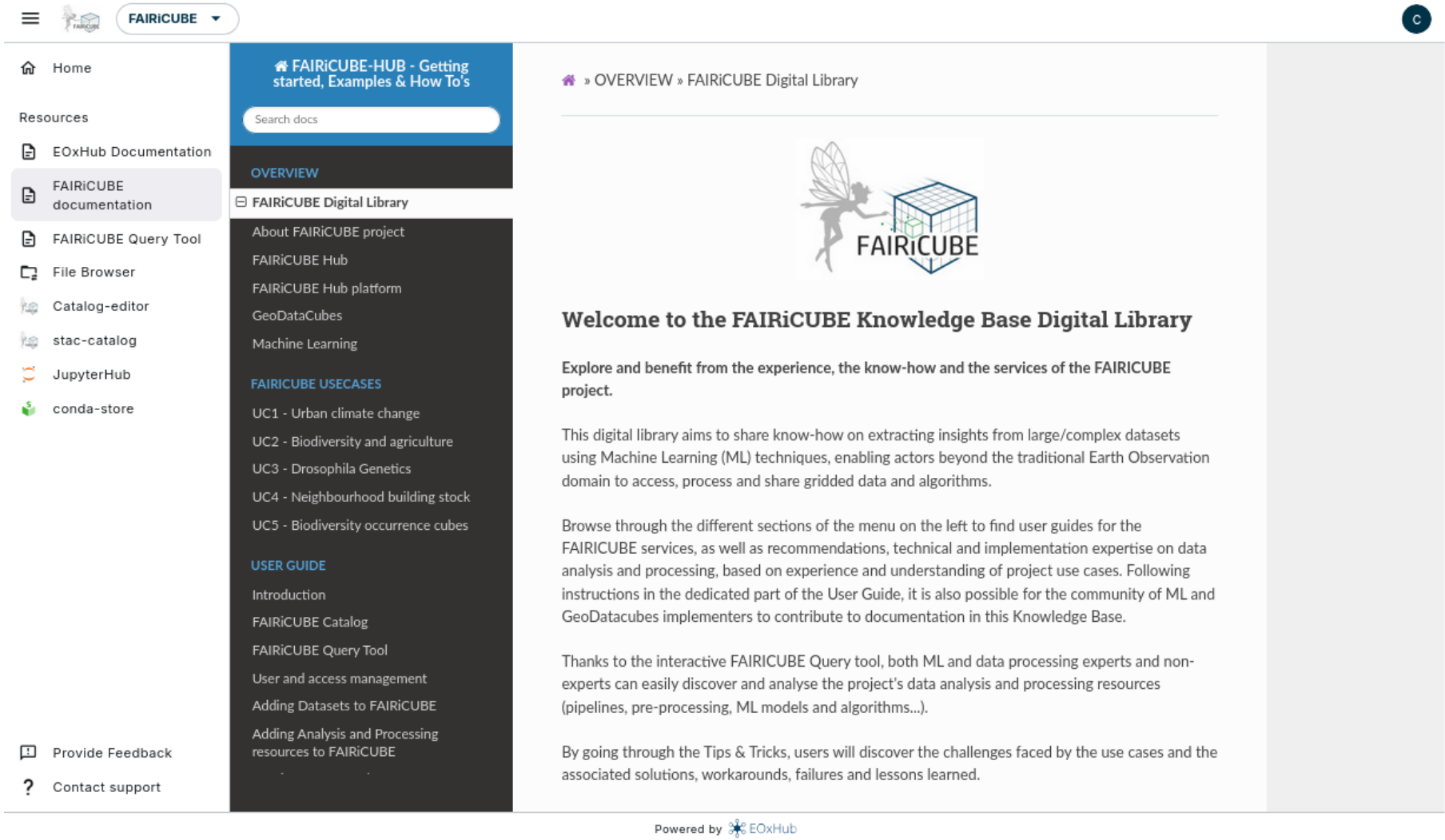
- One well-arranged workspace
- All Apps handled via gateway and inside one namespace
- Making workflows smoother and collaboration more effective
- Enhanced feedback mechanisms for users (incl. screenshots)



https://workspace.fairicube.hub-otc-sc.eox.at/

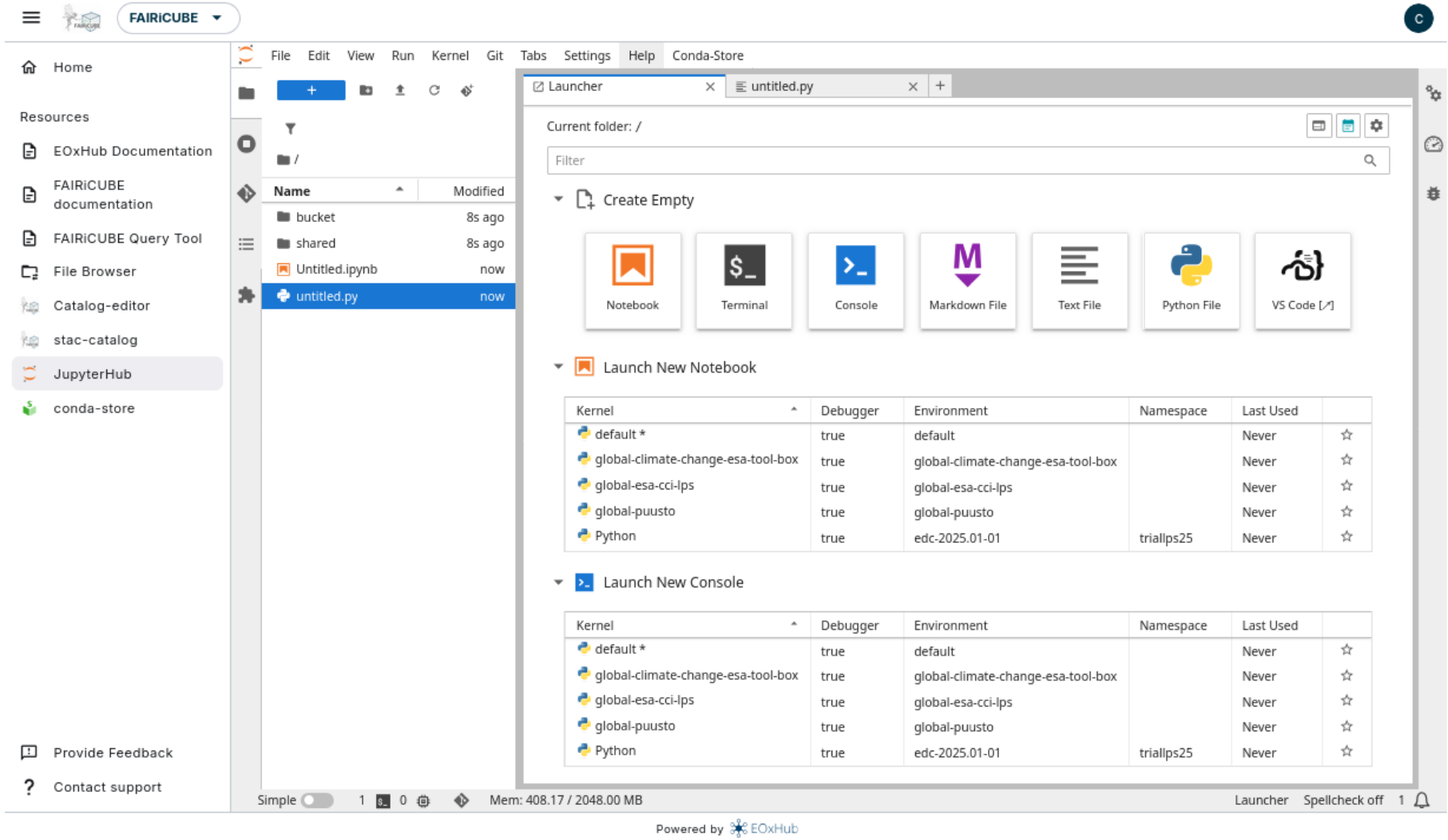
FAIRICUBE LAB - DOCUMENTATION

- Integrated access to the FAIRiCUBE's Digital Library
- Integrated access to Query Tools
- Integrated access to EOxHub documentation



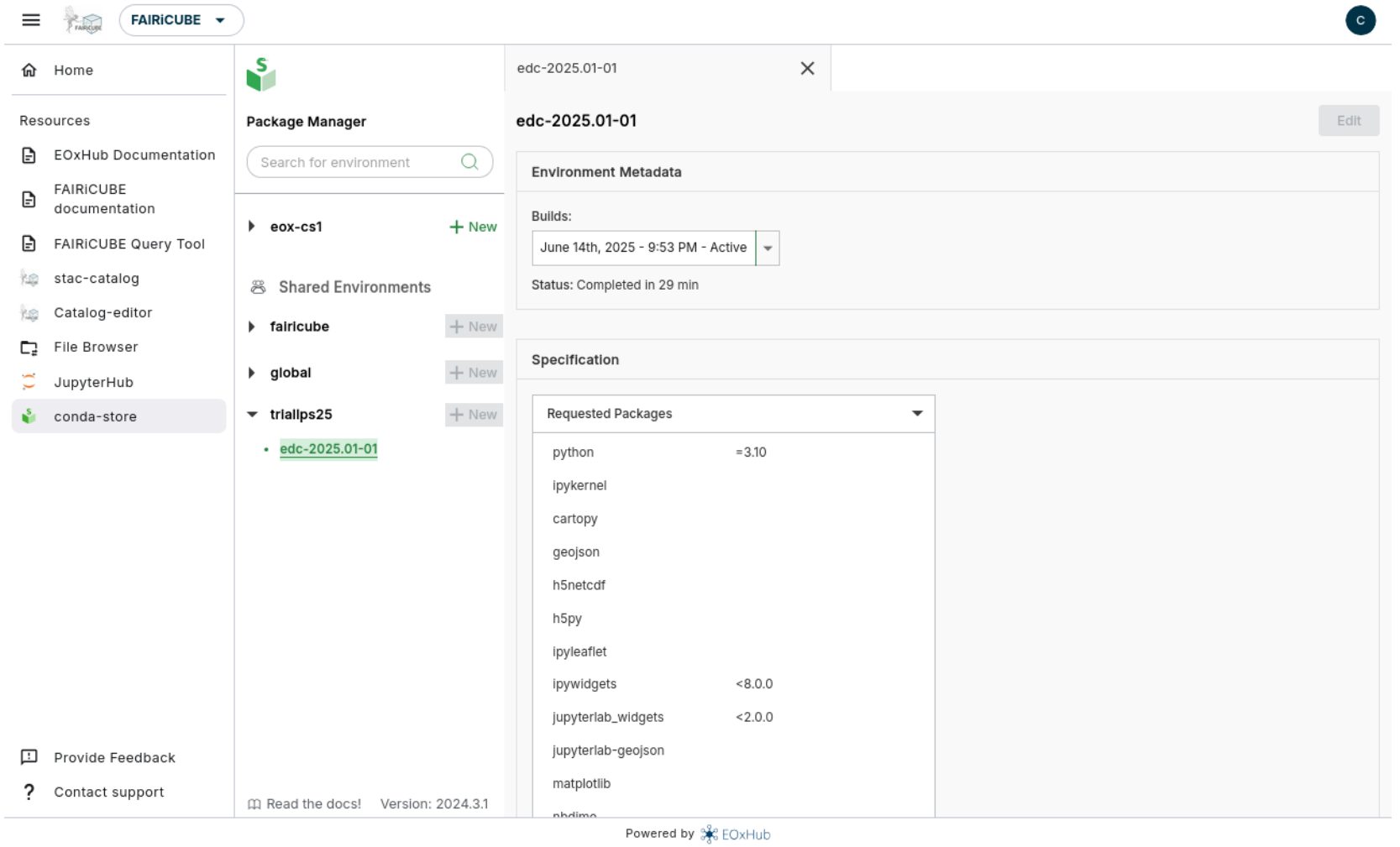
FAIRICUBE LAB – JUPYTER

- Better consolidated Jupyter workspace
- Multiple choices of environments and resources
- Access to Dask gateway
- Enhanced data-space with better sharing possibilities via pre-signed URL generation (single files or folders; enforced via Proxy)



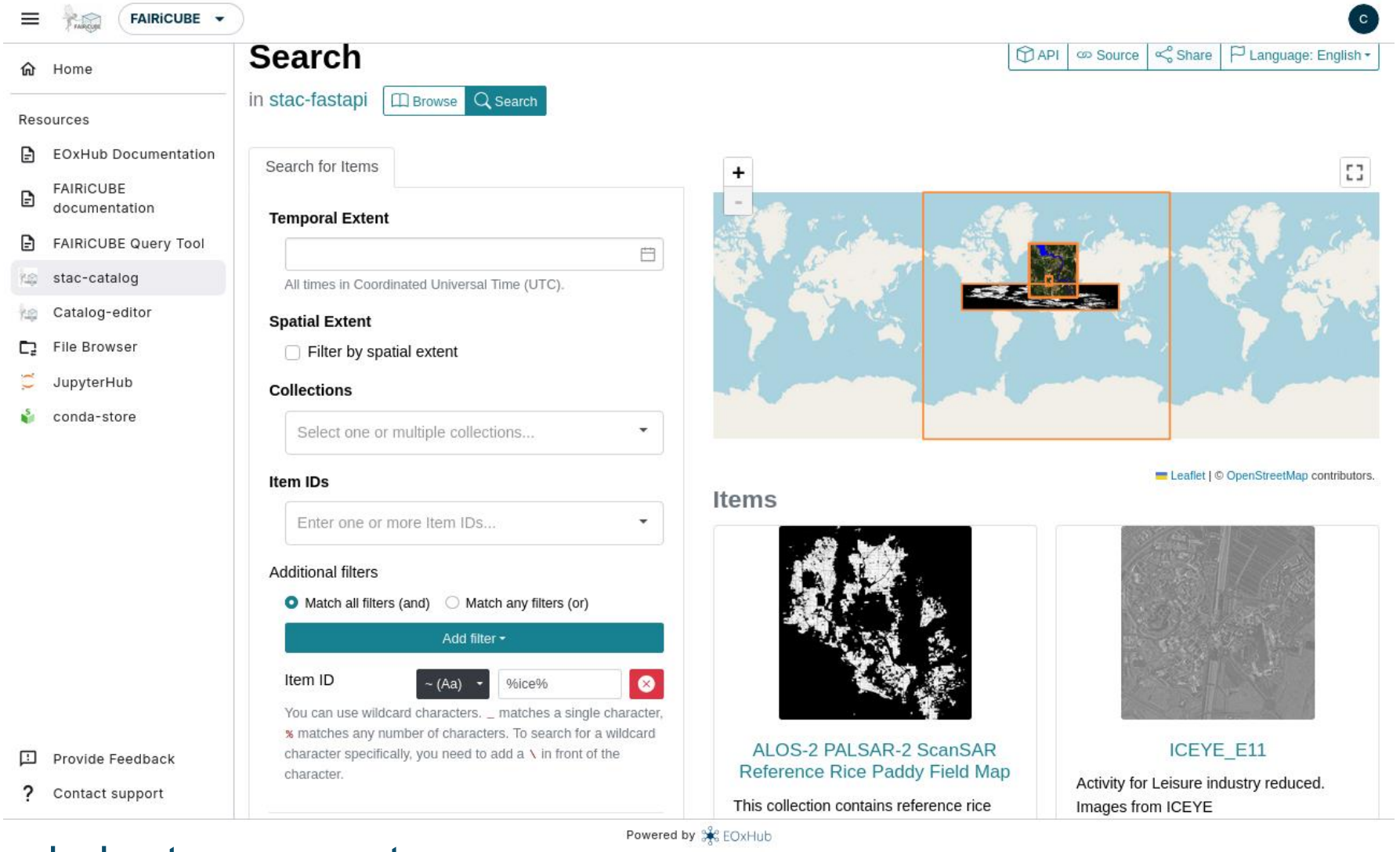
FAIRICUBE LAB – CONDA STORE

- Enhance conda-store interface
- Easier kernel generation by users
- Better kernel management by users
- Easier adaptations of workspace to user's specific needs



STAC BROWSER – SEARCH

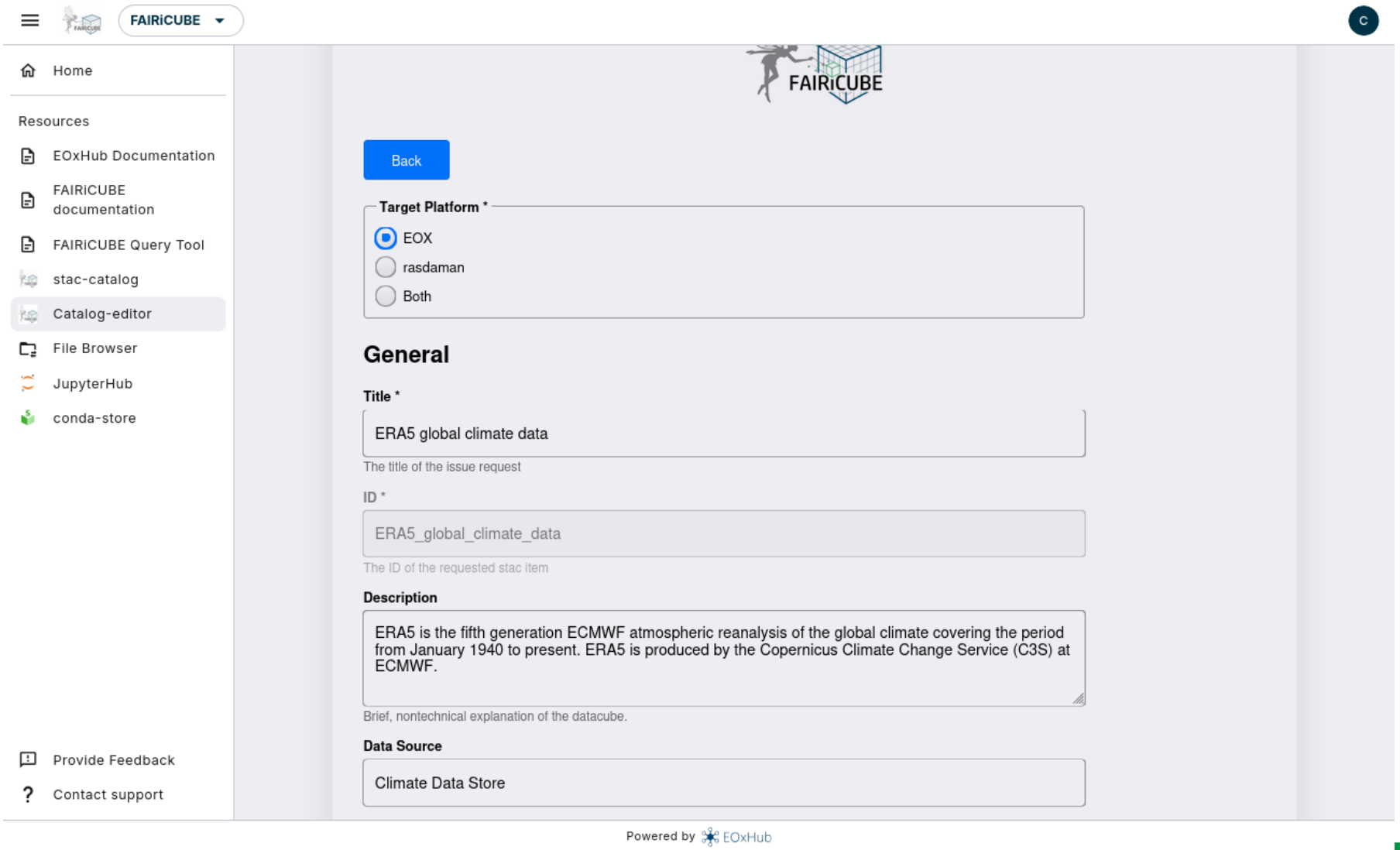
- Integrated access to STAC fastAPI Catalog Browser
- FAIRiCUBE Catalog migrated to version 2
- Advanced Free-text Search function added
- Also accessible directly



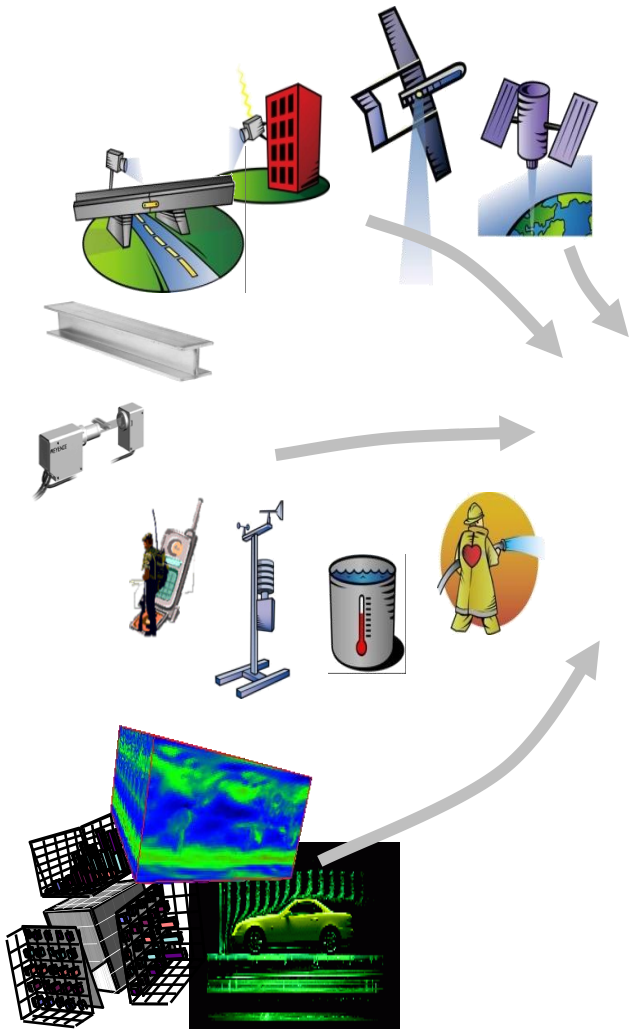
at: http://catalog.fairicube.hub-otc-sc.eox.at

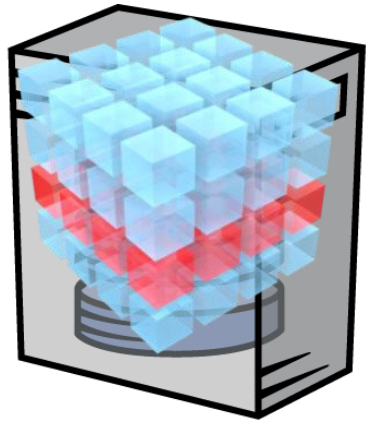
FAIRICUBE STAC CATALOG EDITOR

- Integrated access to the STAC-based Catalog-Editor
- Adapted to the latest FAIRiCUBE proposed STAC extensions and enhancements
- A tool to supply metadata in the correct structure and syntax



RASDAMAN WORKSPACE





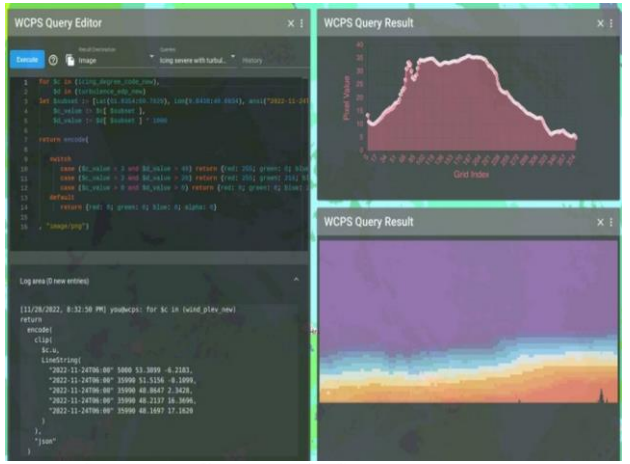
WMS / WMTS

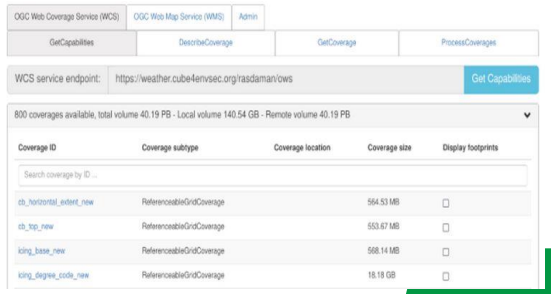
WCPS

openEO

WCS

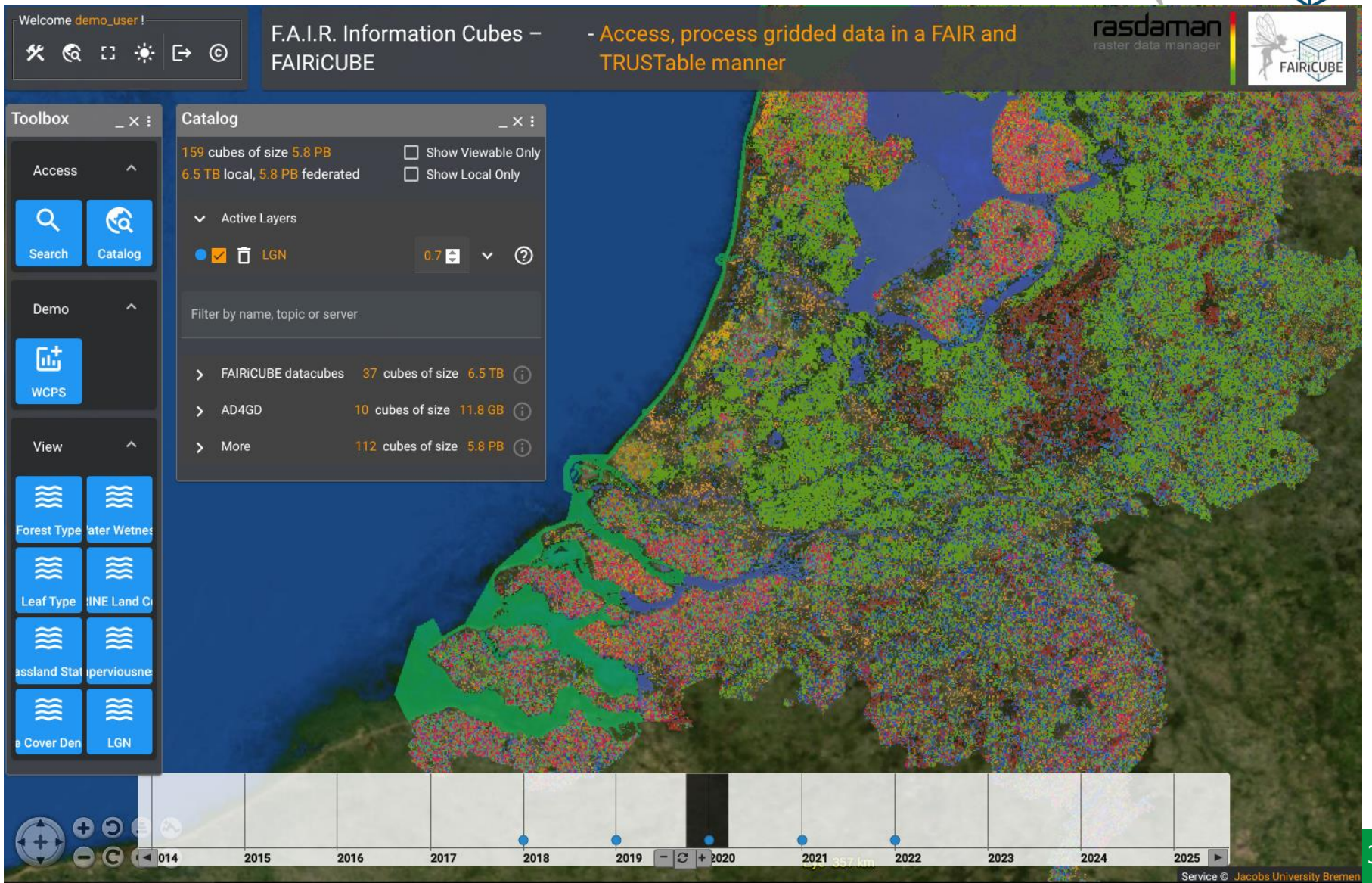
OGC API coverages





DASHBOARD INTERFACE

- LGN datacube - Land Use Database of the Netherlands
- Toolbox (left)
  - Catalog
  - WCPS editor
  - Predefined views
- Temporal navigator (bottom)
- Globe base layer (background)



FAIRICUBE – EVOLUTION

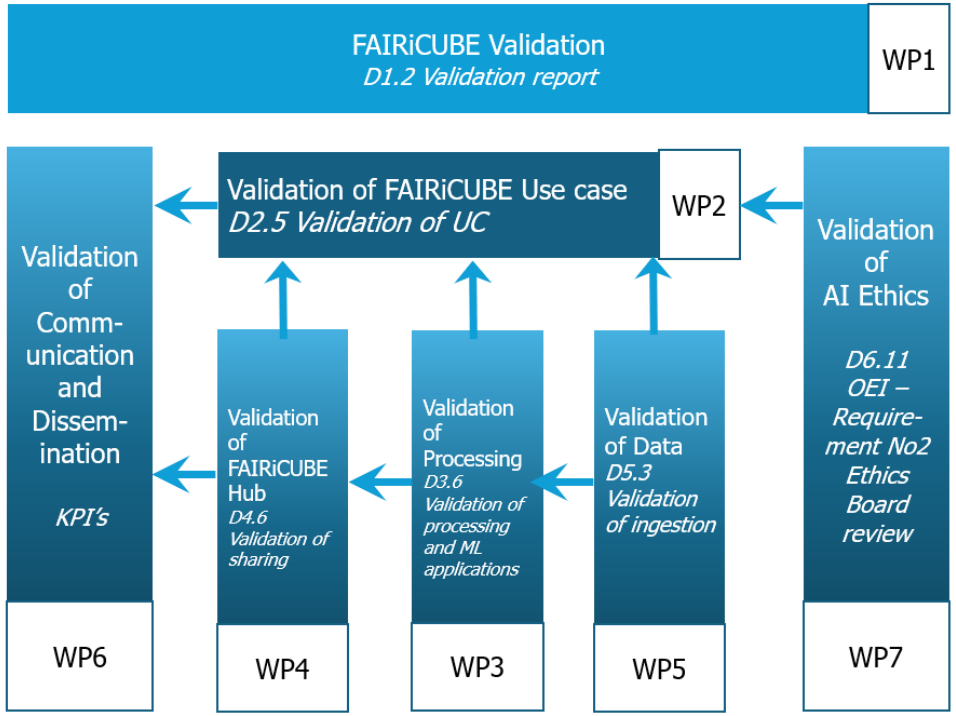
- More seamless access to EO data, applications, and tools
- Improved usability for researchers and developers
- A stronger foundation for FAIRiCUBE's mission to make EO data Findable, Accessible, Interoperable, and Reusable (FAIR)
- Easier user management via Admin console
- Self registration with separate approval step, additionally integrated into a "Slack" workflow
- … and many more Back-end enhancements

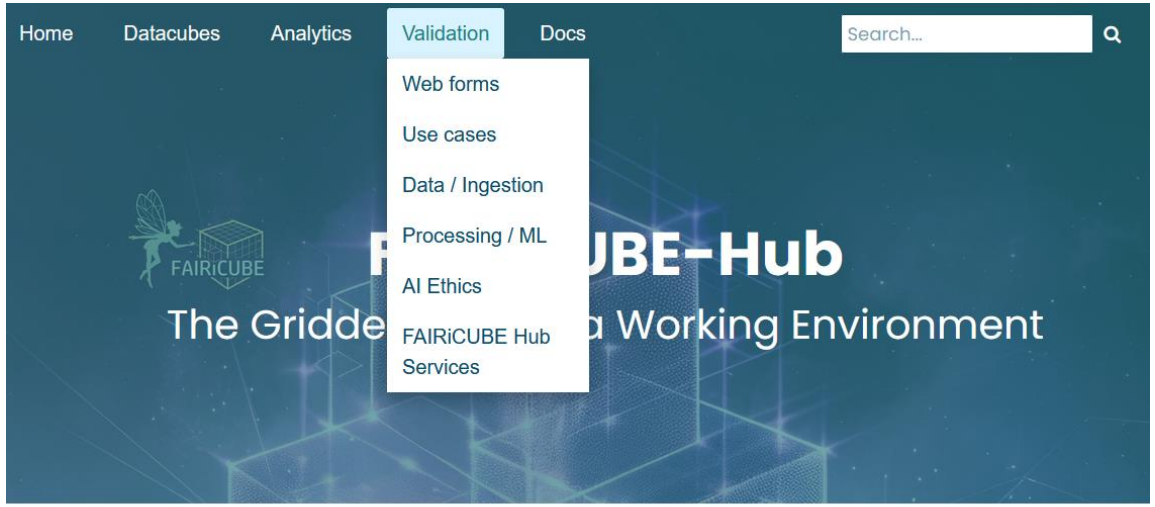
OUTLOOK

- FAIRiCUBE LAB will still be available for 12 months after the project end
  - FAIRiCUBE: Demonstration of results, notebooks and workflows
- The underlying EOxHub services have become a centrally important enabler and will sustainably support various initiatives:
  - EuroDataCube: a marketplace for satellite data, Jupyter notebooks and insights on demand
  - EarthCODE: open-access platform that promotes FAIR Open Science principles and collaboration in Earth System Science research
  - DeepESDL: ESA's Deep Earth System Data Laboratory
  - PolarTEP: Polar Thematic Exploitation Platform
  - GTIF: Austrian Demonstrator GTIF, Cerulean GTIF, Baltic GTIF
  - DESIDE: Supporting policy and decision-makers in decreasing pollution and optimising routes in polar regions

Validation framework

- Holistic validation from start to finish





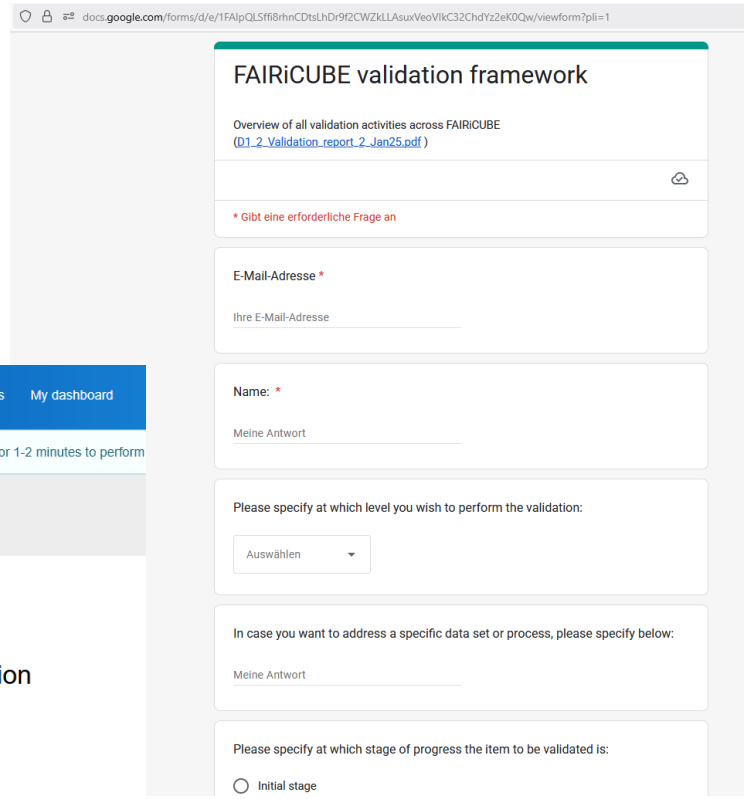
Validation of AI Ethics

The AI ethics assessment in FAIRiCUBE ensures that artificial intelligence is used responsibly and transparently across all Use Cases. The assessment examines how AI models are developed, how transparent their processes are, and whether they avoid bias or unfair outcomes. It also provides practical guidance to ensure that AI is implemented ethically throughout the project and reports on the validation results in a structured and traceable way.

Validation item	Description
Ethics (Trustworthy AI)	
Fundamental rights	How are you dealing with the effect of the application on the rights to safety, health, non-discrimination, and freedom of association?
Privacy and data protection	How are you implementing the GDPR to safeguard the personal data protection rights of those you collect personal data from?

Validation framework

- Guidance and advice (report, deliverables)
- Checklists
- Webform
- Sharing of feedback results



Validation of FAIRiCUBE use cases

Validation of use cases (UCs) in FAIRiCUBE focuses on assessing whether each case is clearly defined, aligned with project goals, and practically applicable. It involves checking the completeness and clarity of UC specifications, evaluating whether the right data and methods are used, and conducting user assessments to ensure the outputs meet real-world needs. Fit-for-purpose checks are applied to confirm that the results are suitable for the intended context. This process also includes documenting findings and communicating them effectively to stakeholders. Importantly, UC validation is closely connected to the validation of processing workflows and platform functionality, ensuring consistency across technical and thematic components.

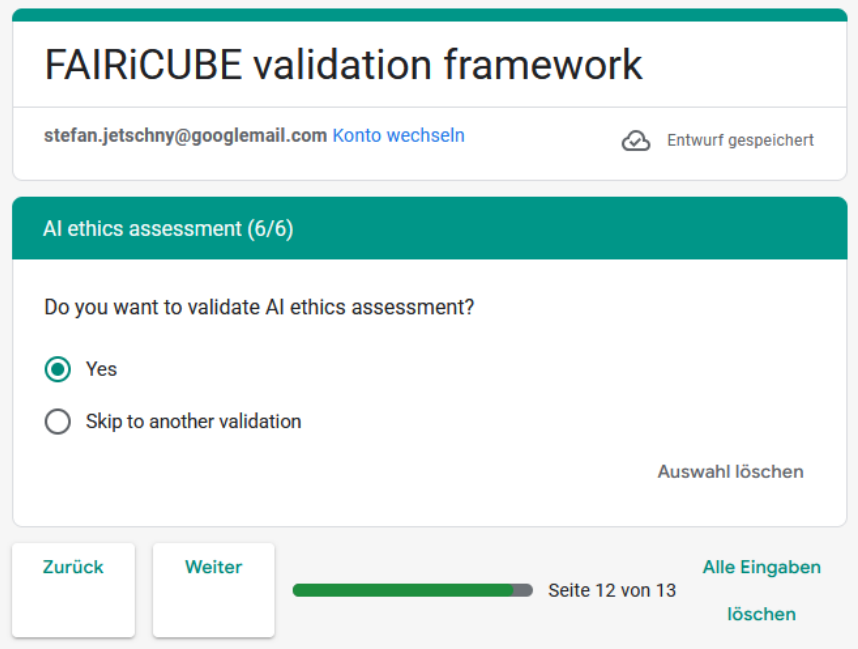
Use Case 1 | Use Case 2 | Use Case 3 | Use Case 4 | Use Case 5

Use Case 1

UC implementation step	Check type	Reference
UC specifications	Clear Goal defined	FAIRiCUBE Digital library – UC1 – Research questions
	Required datasets identified	FAIRiCUBE Digital library – UC1 – Data and ingestion
	Required ML/AI approaches identified	FAIRiCUBE Digital library – UC1 – Processing steps and ML applications
	Workflow designed	FAIRiCUBE Website – UC1 workflow
	Visualisation of outputs designed	UC1 Scrollytelling
User assessment / fitness-for-purpose	Support the users' work	Stakeholder engagement
	Service orientation	
	Reliability	
	Applicability	
	Data systems stability, reliability	

AI ethics

- Obligation from project ethics evaluation
- Renowned experts forming the [AI] Ethics Board
- Starting point: Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment
- FAIRiCUBE : Custom Socio-Technical Scenarios
  - Aim of AI system, actors, context, interaction….
  - Guided walkthrough with pilot use case
  - Generalization and implementation in general validation deliverable
- Generalized implementation in FAIRiCUBE validation framework
  - Validation checklist
  - Part of UC documentation
- Wider Impact : NILU affiliation as **Trustworthy AI Lab**



Input for Green deal data spaces

Kathi Schleidt, Stefan Jetschny

- Meta data
- Validation framework
- AI ethics consideration

INPUT FOR GREEN DEAL DATA SPACES: POLICY BRIEF

- Created in collaboration with our siblings
- Concise, 4 A4 pages (A3 folded)
- Focusing on the topics
  - Data Ingestion
  - Metadata
  - Semantics
  - Data Provision
  - Governance
- Presentation at Living Planet Symposium (LPS)
- doi.org/10.5281/zenodo.15649164

AD4GD FAIRiCUBE USAGE Urban Data Space for Green Deal BIODIVERSITY BUILDING BLOCKS FOR POLICY

POLICY BRIEF

AD4GD BIODIVERSITY BUILDING BLOCKS FOR POLICY FAIRiCUBE USAGE

UNLOCKING THE FULL POTENTIAL OF THE GREEN DEAL DATA SPACE

Making environmental data work for Europe's green transition

Environmental data are essential enablers of the European Green Deal. They support informed decision-making, underpin effective regulation, drive innovation, and empower society to act. Yet today, much of these data remains fragmented, inaccessible, or underutilised, creating significant barriers that limit Europe's progress towards its climate goals.

The Green Deal Data Space (GDDS) is envisioned as a solution to overcome these challenges and foster trusted exchange of public and private data across all areas of the green transition. To be effective, the GDDS must uphold the principles of Findability, Accessibility, Interoperability and Reusability (FAIR). Four projects—AD4GD, B-Cubed, FAIRiCUBE and USAGE—funded under the HORIZON-CL6-2021-GOVERNANCE-01-17 call have explored and documented the requirements for successfully achieving this. The GDDS will be implemented through the SAGE project (The Data Space for a Sustainable Green Europe) and will build on the results of the GREAT initiative (Governance of REsponsible innovATion).

After three years of intensive work, the four projects here join forces to present, in a single comprehensive document, a set of recommendations to guide the successful implementation of the Green Deal Data Space in the years ahead. These are as follows:

- **Strengthen data harmonisation, but stop reinventing the wheel.** Adopt and expand cross-domain standards, based on INSPIRE and other established frameworks, avoiding proprietary formats. Prioritise investment in tools that support data transformation and alignment across existing data structures.
- **Ensure resilient semantic interoperability on top of strong technical foundations.** Promote and enrich comprehensive controlled vocabularies with stable, well-defined concepts under established ontology frameworks, supported by technical infrastructure and governance for adaptability and long-term sustainability.
- **Recognise and resource the effort behind metadata.** Foster interoperability among varied metadata formats to ease the work of producers and users, and expand existing metadata standards to cover critical overlooked elements. Assure resources for metadata creation.
- **Enable data exchange between diverse stakeholders of the GDDS.** Develop effective strategies that motivate open data providers to participate, ensuring a balanced representation of public and commercial data within the data space, and promoting the use of standardised, GDPR-compliant, and federated technologies for data provision.
- **Establish inclusive, participatory and dynamic GDDS governance aligned with the European Green Deal.** Prioritise public interest, long-term sustainability, and adaptability. Leverage existing European data initiatives and state of the art solutions to preserve data sovereignty and security. Provide proper training, tools, and guidance—including real-world examples—to support effective adoption by participants of the data space.

Funded by the European Union

UKRI Innovate UK Knowledge Transfer Partnerships

POLICY BRIEF - DATA HARMONIZATION

Strengthen data harmonisation, but stop reinventing the wheel.

- Investing in **generic cross-domain standard data structures, vocabularies** and representations that operate across a broad spectrum of complexities; avoiding too simplistic formats that ignore semantic links and metadata.
- Promoting **open and well-documented formats**, and ensuring easy transformation towards them, as the use of closed proprietary formats risk vendor lock-in.
- **Building on existing standards and guidelines** for existing data formats, for instance, associating data to the vocabularies already established by INSPIRE across 34 thematic domains — many of which align with the European Green Deal — while continuously identifying gaps and ensuring their sustainable adaptation.
- Investing in **tools that facilitate data transformation**, including libraries and data structure mappings, enabling each community to shift between simpler, purpose-built or domain specific formats and richer, cross-domain generic formats as needed.

POLICY BRIEF - SEMANTIC INTEROPERABILITY

Ensure resilient semantic interoperability on top of strong technical foundations.

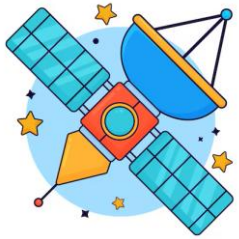
- Promoting means for **linking data and metadata**, including variables, observed properties and units of measure, to stable, persistent and well-defined concepts governed by shared, **linkable, controlled and sustainable vocabularies and ontology frameworks** that ensure robust preservation of meaning, such as the OGC/ISO Observations, Measurements and Samples standard.
- Where applicable, **adopt existing standards** for data structuring and annotation; when strictly necessary, extend them to cover gaps.
- **Enrich internationally agreed and widely adopted vocabularies** — such as the Essential Variables framework — with robust meanings, leveraging their established use within their communities of practice in each domain, e.g. utilising the I-ADOPT Ontology for the semantic representation of variables.

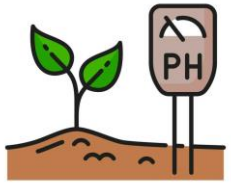
POLICY BRIEF - METADATA

Recognise and resource the effort behind metadata.

- **Expanding existing metadata standards** to include poorly covered critical elements, such as Observations, Measurements & Samples, I-ADOPT, Data Product Specifications, and Data Provenance.
- Identifying and promoting approaches — e.g. **mappings and transformations tools** — to **enhance interoperability** among metadata standards, addressing the needs of commercial and open data alike.
- Ensuring that European research funding explicitly acknowledges the **time and resources necessary for metadata creation and maintenance**, and making it mandatory for successful project implementation.
- For European research projects, incorporate **requirements for the creation and maintenance of a standardised metadata catalogue** into the data management plan.

Policy Brief - Meta Data





- **Satellite/EO**
  - Standardized information on satellite instrumentation
  - Deficits when describing derived products
    - Observable properties and processing relegated to prose
    - Poor provenance information
- **Terrestrial *in situ***
  - Long history of measurement meta-information, e.g.
    - OGC/ISO Observations, Measurements and Samples
    - PROV-O Provenance Ontology
    - I-ADOPT Observable Property disaggregation

- GDDS requires data from both worlds. Work still needs to be done to merge the strengths from both worlds.

POLICY BRIEF - DATA EXCHANGE

Enable data exchange between diverse stakeholders of the GDDS

- Adopting and promoting **open, easy to use technologies** that support seamless integration of both open and protected data.
- Ensuring the **integration of diverse in-use APIs** from data providers and user communities into the data space environment, accompanied by demonstrators and clear guidelines.
- Promoting the adoption of **standardised APIs** to facilitate smoother uptake, for instance the OGC APIs for geospatial assets.
- Offering **conversion tools**, like facades and brokers, and clear specifications to enable interoperability between APIs and data formats.

POLICY BRIEF - GOVERNANCE

Establish inclusive, participatory and dynamic GDDS governance aligned with the European Green Deal.

- **Integrating existing open-data platforms** such as the Copernicus Open Access Hub, the European Data Portal, and the INSPIRE Geoportal within the GDDS.
- Developing effective strategies to **attract open data providers** to join the GDDS, ensuring that the dataspace does not become exclusively commercial or paywalled.
- **Mandating the citation of data** in publicly funded academic publications to increase the visibility and recognition of open data, encouraging data providers to share their datasets more openly.

FINAL WORDS

- *"Every data scientist working with EO data can potentially benefit from FAIRiCUBE"*
  S. Jetschny, 2022

- Concrete technical solutions, guidelines and examples to:
  - How can we scale up work from local to cloud resources?
  - How can we include data from different sources?
  - How can we exploit ML techniques?
  - How can we make data science work as F.A.I.R. as possible?
  - How can we validate use cases from start to finish?
  - **How can we bring technology, EO data and scientists together?**

Many thanks for your attention!

This project has received funding from the European Union's horizon 2020 research and innovation programme under grant agreement no 101059238

… and to this great consortium!