Erasmus Data Collaboratory

Erasmus Centre for Data Analytics

AI, Data & Digitalisation

EnergySHR, a Data Sharing & Collaboration Platform

Erasmus University Rotterdam

The Erasmus data Summit was founded by

DATA REVOLUTION FOR A SUSTAINABLE WORLD

Erasmus Centre for Data Analytics

Flagship centre and community
for cross-disciplinary insight
on the societal impact of Data,
Artificial Intelligence, Digitalization
and Immersive Technologies
at the Erasmus University Rotterdam.

Erasmus Data Collaboratory | House of AI

Convergence initiative to foster an inclusive, open experimenting environment for data and AI, where people can collectively think, work, innovate, and learn. The EDC | House of AI is facilitated by the Erasmus Centre for Data Analytics

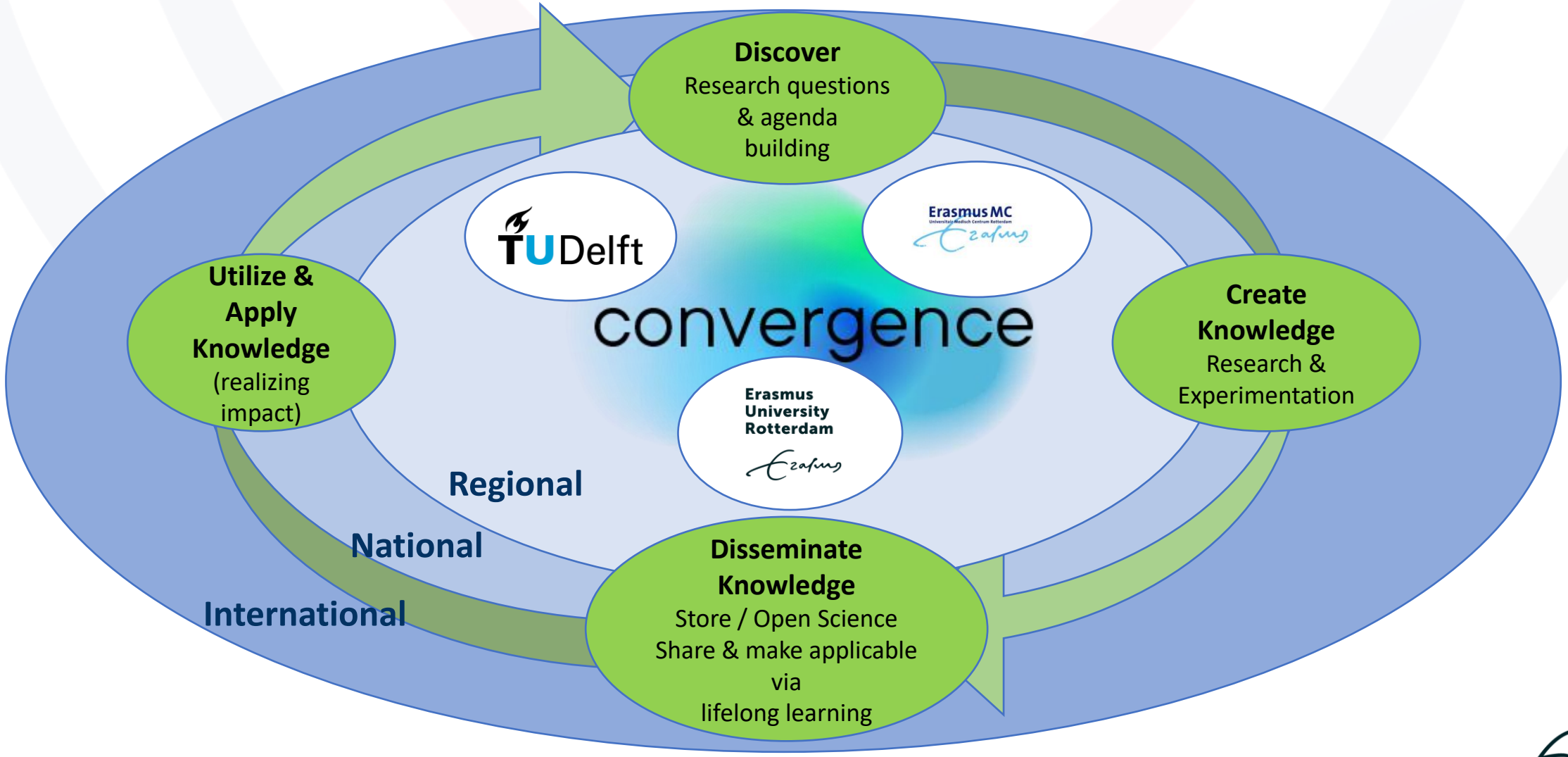
Erasmus

Convergence

The urgent and complex societal challenges of our time call for convergence; transcending boundaries between institutes and disciplines to create new perspectives and solutions. For this reason, TU Delft, Erasmus MC and Erasmus University Rotterdam have joined forces in Convergence.

TU Delft · Erasmus MC · Erasmus University Rotterdam

Building Learning Communities & Innovation Ecosystems



INNOVATING RESEARCH THROUGH OPEN DATA: BUILDING THE COLLABORATION PLATFORM OF THE FUTURE

Findable, for humans and machines; metadata & DOI's

Accessible, within well defined conditions; metadata access

Interoperable, open standards and structures, compatibility

Reusable, with clear documentation, licensing and context for easy use in research

Digital Object Identifier (DOI)

Erasmus University Rotterdam

Unique, permanent alphanumeric string assigned to digital objects, such as journal articles, books, datasets, and research reports, to ensure they can be reliably identified and accessed on the internet. It serves as a persistent link to the object, even if its location (URL) changes over time.

Key features

- **Uniqueness:** Each DOI is unique to the object it identifies.
- **Persistence:** The DOI remains constant, even if the object's URL changes. This ensures long-term access through updates in the DOI database.
- **Structure:** A DOI consists of two parts:
  - **Prefix:** Always starts with "10" and identifies the publisher or organization.
  - **Suffix:** A unique string assigned by the publisher to identify the specific object (e.g., `10.1016/j.addbeh.2008.11.004`) -> https://doi.org/10.1016/j.addbeh.2008.11.004

ELSEVIER

Addictive Behaviors

Addictive Behaviors

Volume 34, Issue 3, March 2009, Pages 281-286

Testing Social Cognitive Theory as a theoretical framework to predict smoking relapse among daily smoking adolescents ☆

Rinka M.P. Van Zundert, Linda M. Nijhof, Rutger C.M.E. Engels

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.addbeh.2008.11.004

Get rights and content

Full text access

Erasmus University Rotterdam

Erasmus Data Collaboratory

Data Sharing 1985 ☺

Erasmus University Rotterdam

Erasmus Data Collaboratory

Data Sharing 1987

Erasmus University Rotterdam

UCI Machine Learning Repository
Center for Machine Learning and Intelligent Systems

About Citation Policy Donate a Data Set Contact

Search
Repository Web
Google

View ALL Data Sets

Welcome to the UC Irvine Machine Learning Repository!

We currently maintain 622 data sets as a service to the machine learning community. You may **view all data sets** through our searchable interface. For a general overview of the Repository, please visit our About page. For information about citing data sets in publications, please read our citation policy. If you wish to donate a data set, please consult our donation policy. For any other questions, feel free to contact the Repository librarians.

Supported By: In Collaboration With: Rexa.info

Latest News:

- **09-24-2018:** Welcome to the new Repository admins Dheeru Dua and Efi Karra Taniskidou!
- **04-04-2013:** Welcome to the new Repository admins Kevin Bache and Moshe Lichman!
- **03-01-2010:** Note from donor regarding Netflix data
- **10-16-2009:** Two new data sets have been added.
- **09-14-2009:** Several data sets have been added.
- **03-24-2008:** New data sets have been added!
- **06-25-2007:** Two new data sets have been added: UJI Pen Characters, MAGIC Gamma Telescope

Featured Data Set: Statlog (Landsat Satellite)

Task: Classification
Data Type: Multivariate
 # Attributes: 36
 # Instances: 6435

Newest Data Sets:

- **06-05-2021:** Average Localization Error (ALE) in sensor node localization process in WSNs
- **05-25-2021:** 9mers from cullpdb
- **05-18-2021:** TamilSentiMix
- **05-02-2021:** Accelerometer
- **04-21-2021:** Synchronous Machine Data Set
- **04-21-2021:** Synchronous Machine Data Set

Most Popular Data Sets (hits since 2007):

- **4645978:** Iris
- **2466800:** Adult
- **2001229:** Dry Bean Dataset
- **1908607:** Wine
- **1855217:** Wine Quality
- **1852257:** Heart Disease

UCI Machine Learning Repository
Center for Machine Learning and Intelligent Systems

Iris Data Set

Download: Data Folder, Data Set Description

Abstract: Famous database; from Fisher, 1936

Data Set Characteristics:	Multivariate	**Number of Instances:**	150	**Area:**	Life
Attribute Characteristics:	Real	**Number of Attributes:**	4	**Date Donated**	1988-07-01
Associated Tasks:	Classification	**Missing Values?**	No	**Number of Web Hits:**	4659866

Source:

Creator:

R.A. Fisher

Erasmus Data Collaboratory

Data Sharing 2010

Erasmus University Rotterdam

kaggle

About us | How it works | Find a competition | Host a competition | Blog | Help

is a platform for data prediction competitions that allows organizations to post their data and have it scrutinized by the world's best data scientists. See how it works.

CURRENT COMPETITIONS

Hearst Challenge

The objective of this competition is to predict the optimal number and type of newspapers that should be placed on different news stands. **This competition is being powered by Kaggle but run on a standalone website.**

Ends in **5 days** | **58** teams | **$25,000.00**

IJCNN Social Network Challenge

This competition requires participants to predict edges in an online social network. The winner will receive free registration and the opportunity to present their solution at IJCNN 2011.

Ends in **6 weeks** | **38** teams | **USD 950.00**

R Package Recommendation Engine

The aim of this competition is to develop a recommendation engine for R libraries (or packages). (R is opensource statistics software.)

Ends in **10 weeks** | **28** teams | **USD 150.00**

Get started now to create or enter competitions:

Sign up

or just join our mailing list

RETURNING USER SIGN IN

Username

Login | Remember me?

Forgot your Username/ Password?

Current Competitions

Hearst Challenge
5 days | 58 teams | $25,000

IJCNN Social Network Challenge
6 weeks | 38 teams | $950

R Package Recommendation Engine

Erasmus Data Collaboratory

Data Sharing 2025

Erasmus University Rotterdam

Erasmus Data Collaboratory

Why Build a New One?

Erasmus University Rotterdam

ENERGY DATA SHARING HOUSE

Erasmus Data Collaboratory I House of AI
Erasmus Center for Data Analytics
Erasmus University Rotterdam

Erasmus Data Collaboratory | House of AI

convergence

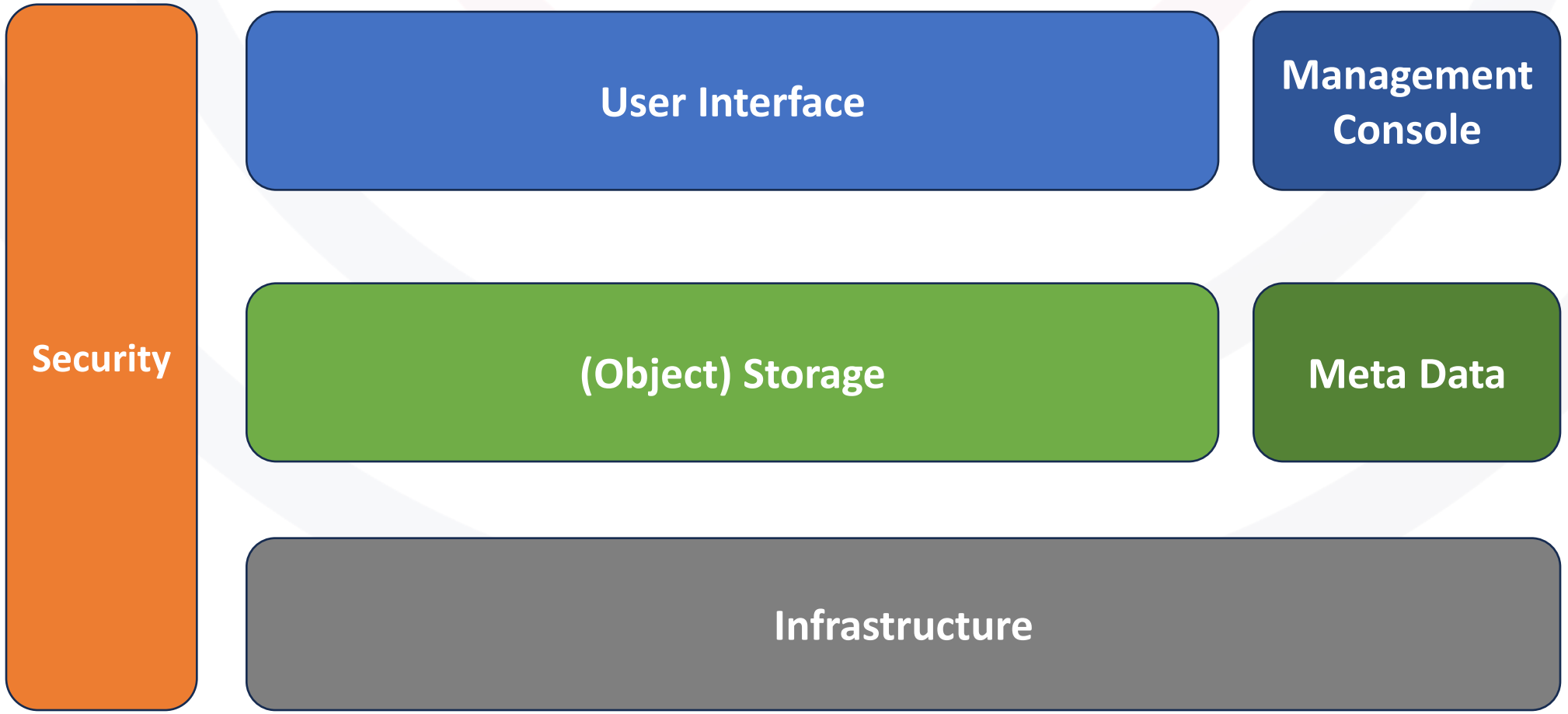
High Level Requirements

Erasmus University Rotterdam

The EnergySHR platform is designed to facilitate the sharing, discovery, and management of datasets and code, with a focus on collaboration, data integrity, and access control. Key functionalities include:

- User Authentication and Access Control: Secure login and permissions management to protect sensitive data.
- Dataset and Code Publication: Tools for publishing datasets and code with detailed metadata and licensing options.
- Search and Discovery: Advanced search features for finding relevant datasets and code efficiently.
- Version Control: Support for tracking changes in datasets and code repositories.
- Collaboration Features: Platforms for discussion, feedback, and collaborative projects.
- Notification System: Alerts for updates or changes to datasets of interest.
- Automatic Updates: Synchronization of derivative datasets with upstream data changes

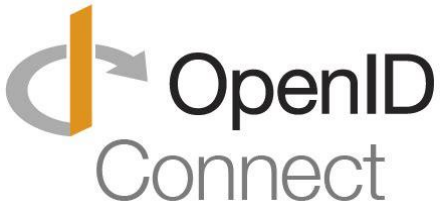
High Level Design



Selecting Software Components

Erasmus University Rotterdam

- Open Source
- Open Standards
- Python based
- S3 compatible



Erasmus Data Collaboratory

Why Django?

Erasmus University Rotterdam

1. **Speed & Efficiency**: Rapid development with built-in tools.
2. **Security**: Robust protection against common web vulnerabilities.
3. **Scalability**: Handles high traffic and large applications.
4. **Documentation & Community**: Excellent resources and support.
5. **Versatility**: Suitable for diverse projects and applications.

Erasmus University Rotterdam

S3-compliant storage

Ease of use: direct data access

Flexible: wide support in tools, vendors

Data variety: structured & unstructured

Scalable: virtually limitless

Cost efficient: on premise; no vendor lock in

S3 Options

Erasmus University Rotterdam

Open-IO

MicroCeph

MINIO

SWIFT

an OpenStack Community Project

GLUSTER

ceph

Amazon S3

leafcloud

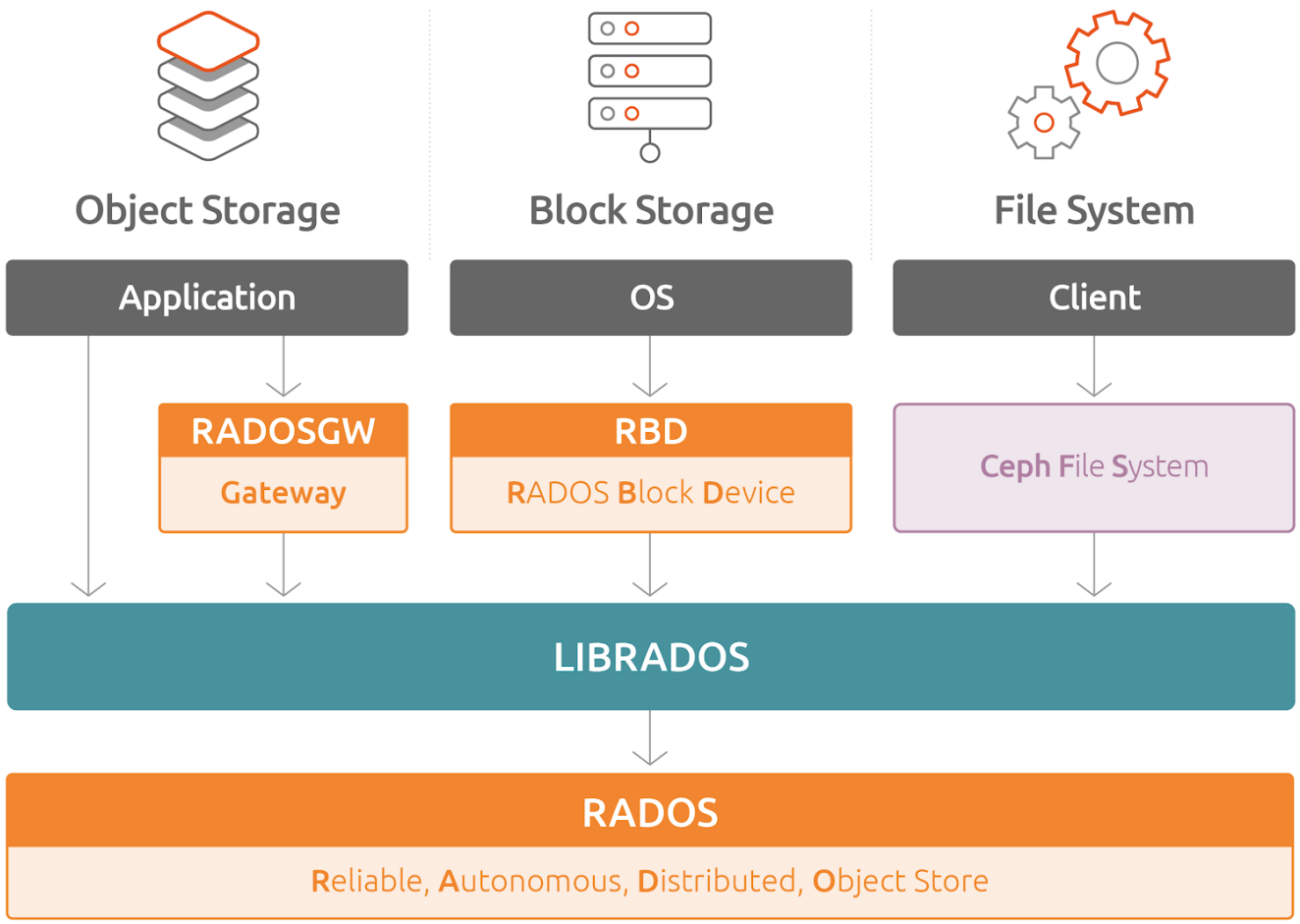
HETZNER

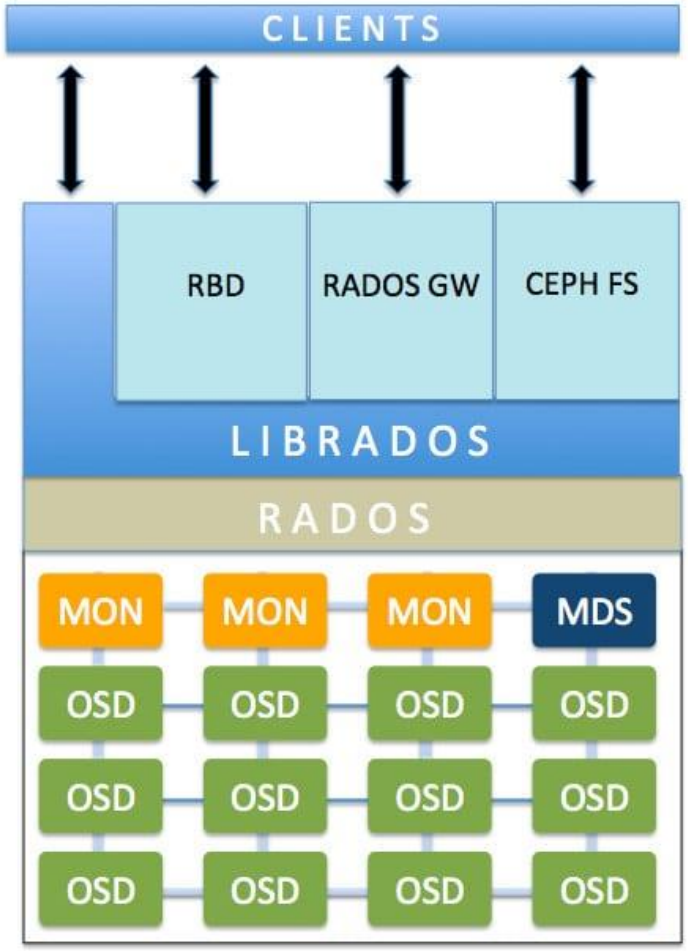
OVHcloud

Microsoft Azure
Blob Storage

Google Cloud Storage

CEPH Overview





Best Ceph Tutorial Ever: https://youtu.be/jJrnJ9rj6fs?si=C94VJv5h8UGSzBV2

Erasmus University Rotterdam
NASA
T-Mobile
CERN
OVHcloud
accenture
NOKIA
intel
Erasmus Data Collaboratory

Ceph in Proxmox

Erasmus University Rotterdam

Health

Status

HEALTH_OK

Ceph Version: 19.2.0

Summary

No Warnings/Errors

Status

OSDs

	In	Out
Up	12	0
Down	0	0

Total: 12

active+clean:

Name	Class	OSD Type	Status	Version	weight	reweight	Used (%)	Total	Apply/Commit Latency (ms)	PGs
default										
mini6				19.2.0						
osd.11	nvme	bluestore	up / in	19.2.0	3.2715	1.00	0.22	3.27 TiB	3 / 3	33
osd.10	nvme	bluestore	up / in	19.2.0	3.2715	1.00	0.17	3.27 TiB	4 / 4	26
mini5				19.2.0						
osd.9	nvme	bluestore	up / in	19.2.0	3.2715	1.00	0.19	3.27 TiB	4 / 4	35
osd.8	nvme	bluestore	up / in	19.2.0	3.2715	1.00	0.34	3.27 TiB	4 / 4	37
mini4				19.2.0						
osd.7	nvme	bluestore	up / in	19.2.0	3.2715	1.00	0.22	3.27 TiB	4 / 4	37
osd.6	nvme	bluestore	up / in	19.2.0	3.2715	1.00	0.33	3.27 TiB	8 / 8	36
mini3				19.2.0						
osd.5	nvme	bluestore	up / in	19.2.0	3.2715	1.00	0.18	3.27 TiB	2 / 2	17
osd.4	nvme	bluestore	up / in	19.2.0	3.2715	1.00	0.28	3.27 TiB	3 / 3	40
mini2				19.2.0						
osd.3	nvme	bluestore	up / in	19.2.0	3.2715	1.00	0.21	3.27 TiB	5 / 5	35
osd.2	nvme	bluestore	up / in	19.2.0	3.2715	1.00	0.27	3.27 TiB	3 / 3	34
mini1				19.2.0						
osd.1	nvme	bluestore	up / in	19.2.0	3.2715	1.00	0.27	3.27 TiB	2 / 2	29
osd.0	nvme	bluestore	up / in	19.2.0	3.2715	1.00	0.12	3.27 TiB	4 / 4	28

Details: OSD 7

Reload

General | Network | Devices

Device	Type	Physical Device	Size
block	ssd	nvme1n1	3.27 TiB

Volume Details for block

Creation time	2025-01-29 12:01:21 +0100
LV Name	osd-block-cdeb74ef-803f-4afa-bb19-8e11266747f6
LV Path	/dev/ceph-209f4379-e3b6-4663-8acb-a2c8cd6fabcc/osd-block-cdeb…
LV UUID	kWsGjj-7mGa-AzkO-6B3W-y4WA-rrpu-tkjFlt
VG Name	ceph-209f4379-e3b6-4663-8acb-a2c8cd6fabcc

Services

Monitors

mini1: ✔ | mini3: ✔ | mini5: ✔

Managers

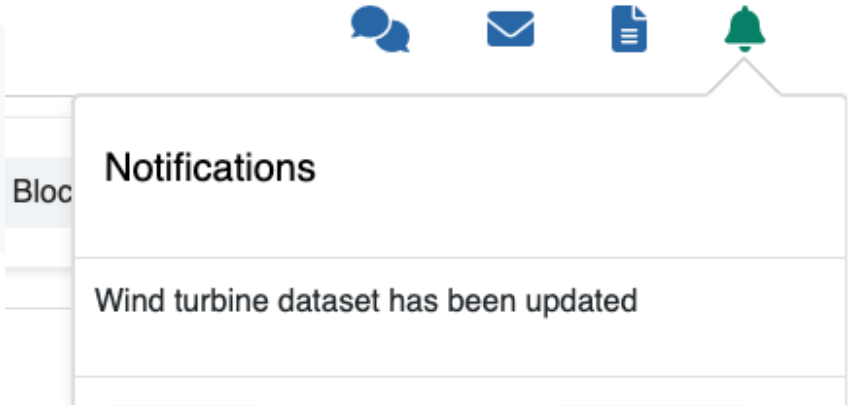
mini2: ✔ | mini3: ✔ | mini4: ✔

Meta Data Servers

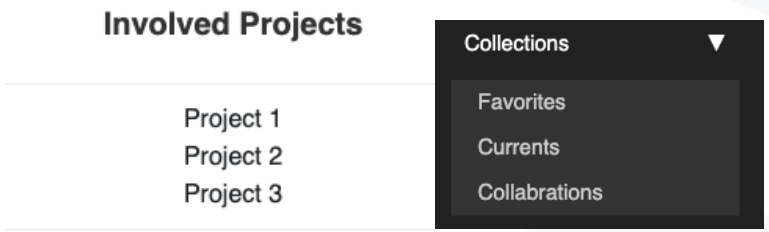
mini1: ✔ | mini2: ✔ | mini4: ✔

Erasmus Data Collaboratory

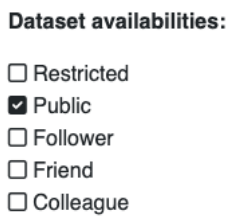
Notable EnergySHR Features



Integrated mail, messaging, discussion & notification system

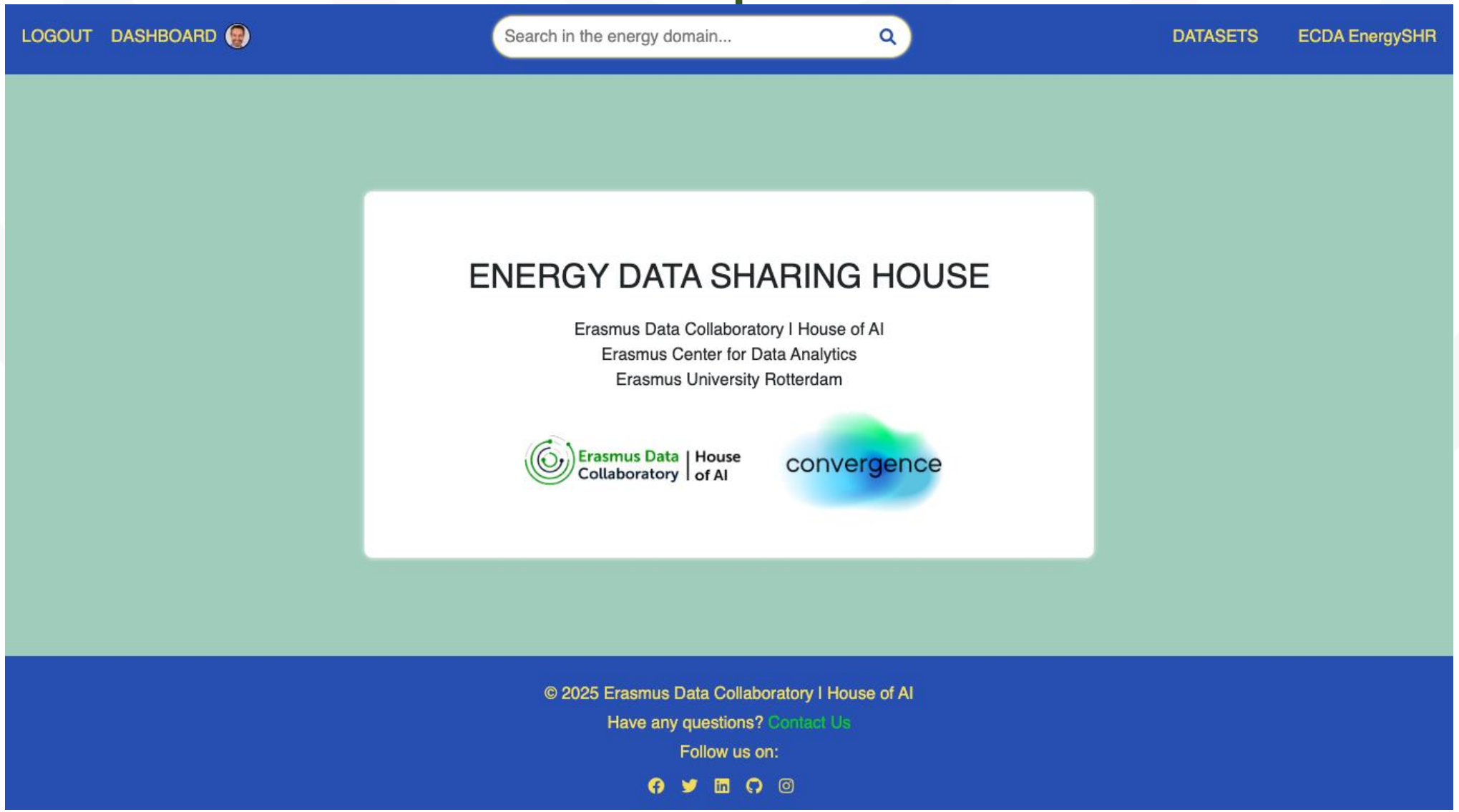


Projects (collaborations) & Collections (Datasets, Code, Doc)

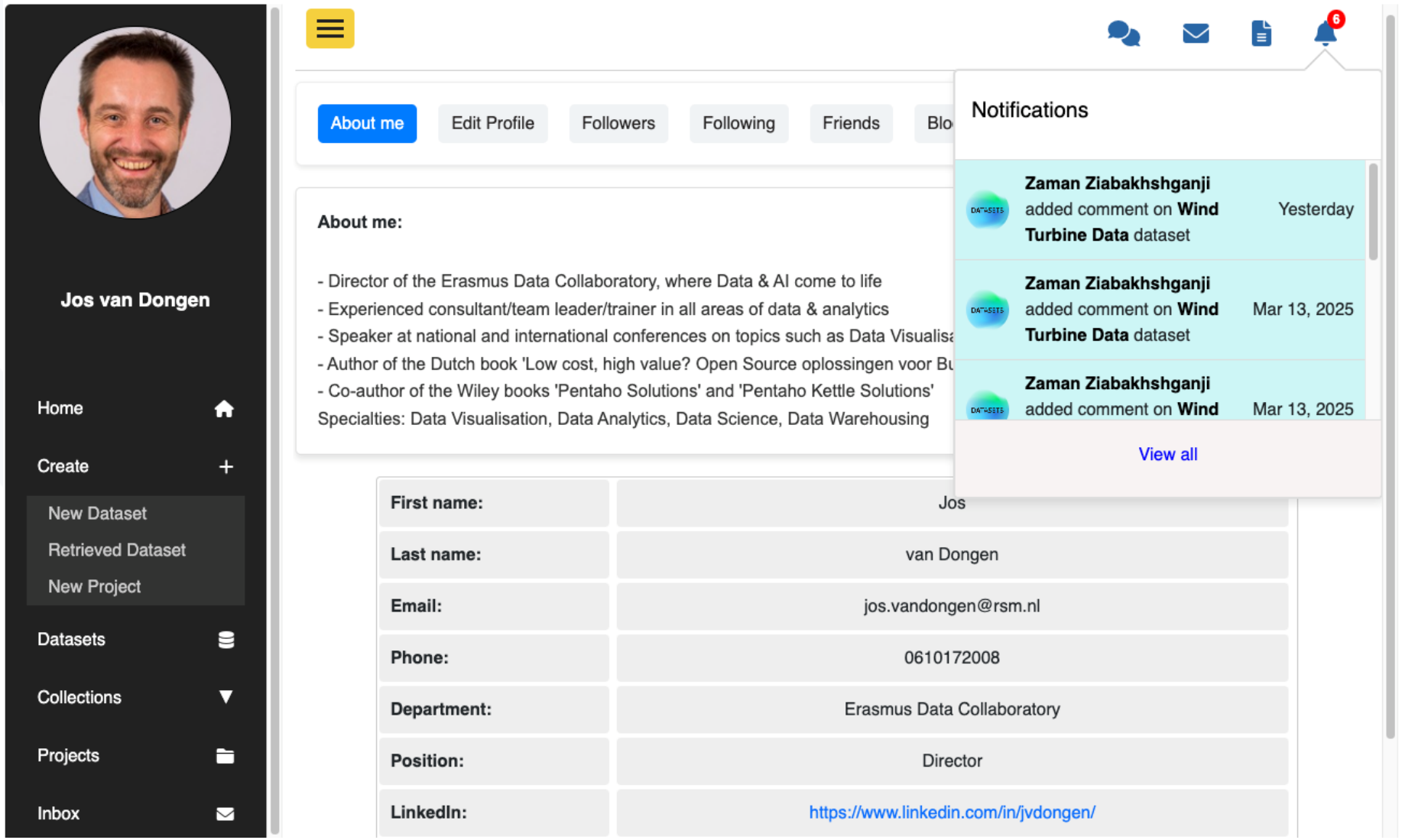


Fine grained sharing & authorization options

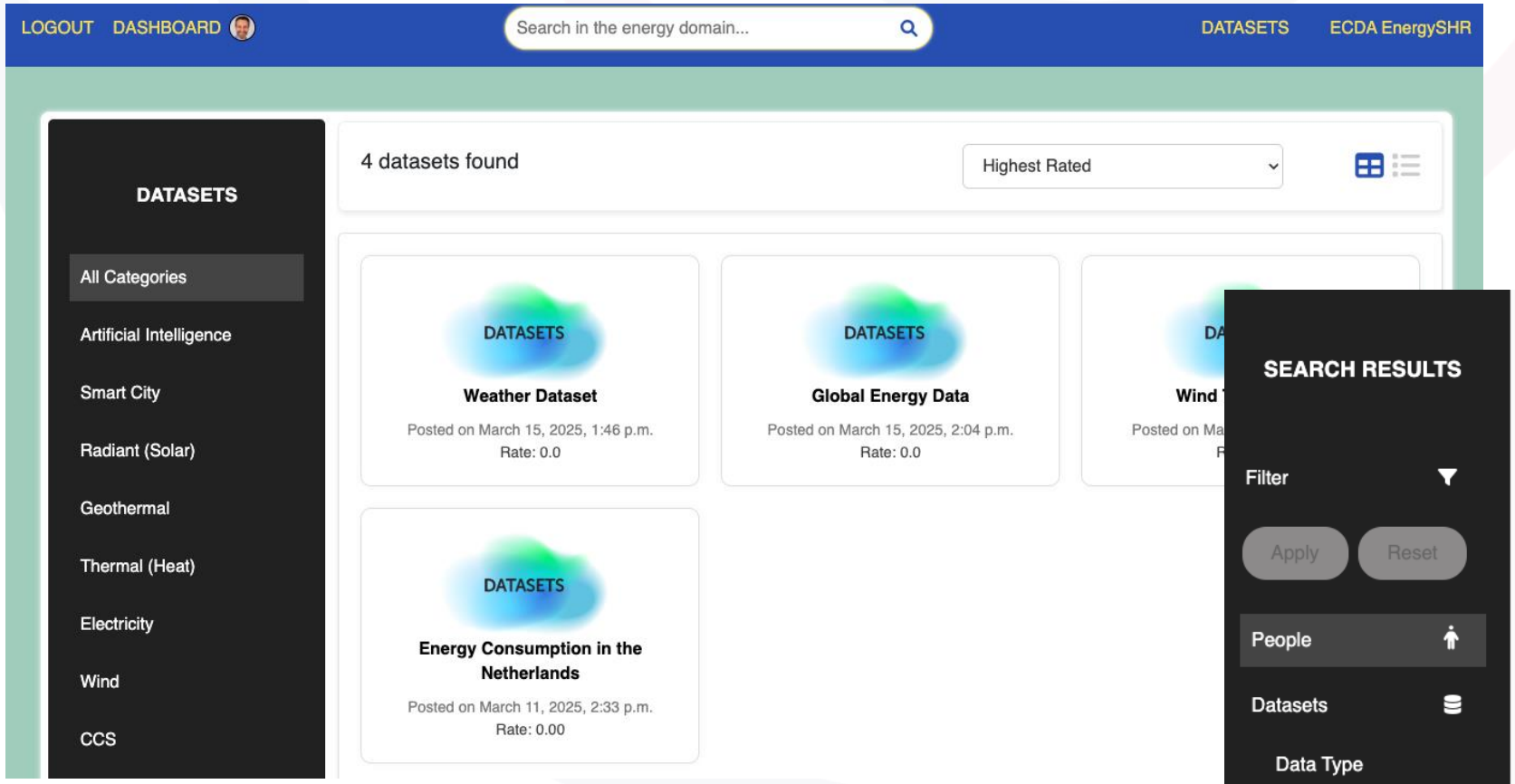
Simple UI

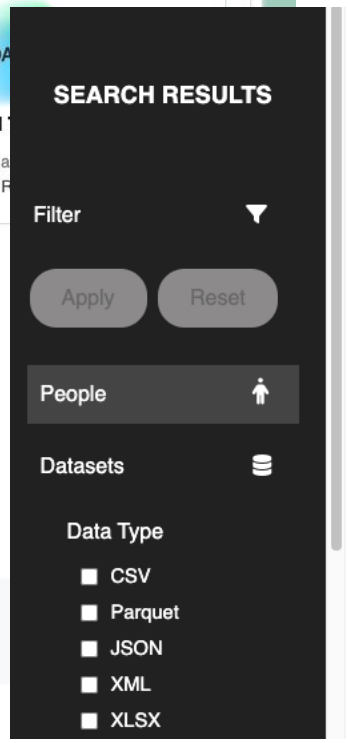

LOGOUT
DASHBOARD
Search in the energy domain...
DATASETS
ECDA EnergySHR
ENERGY DATA SHARING HOUSE
Erasmus Data Collaboratory I House of AI
Erasmus Center for Data Analytics
Erasmus University Rotterdam
Erasmus Data Collaboratory | House of AI
convergence
© 2025 Erasmus Data Collaboratory I House of AI
Have any questions? Contact Us
Follow us on:

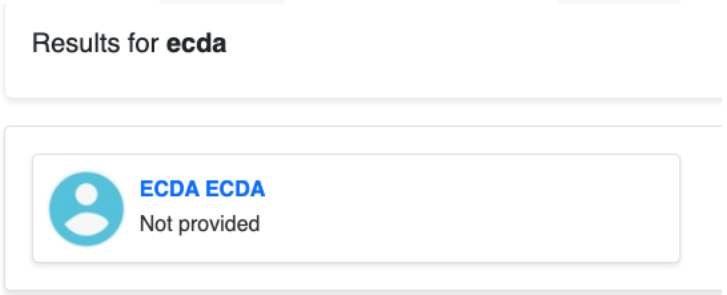
Simple UI: User Dashboard



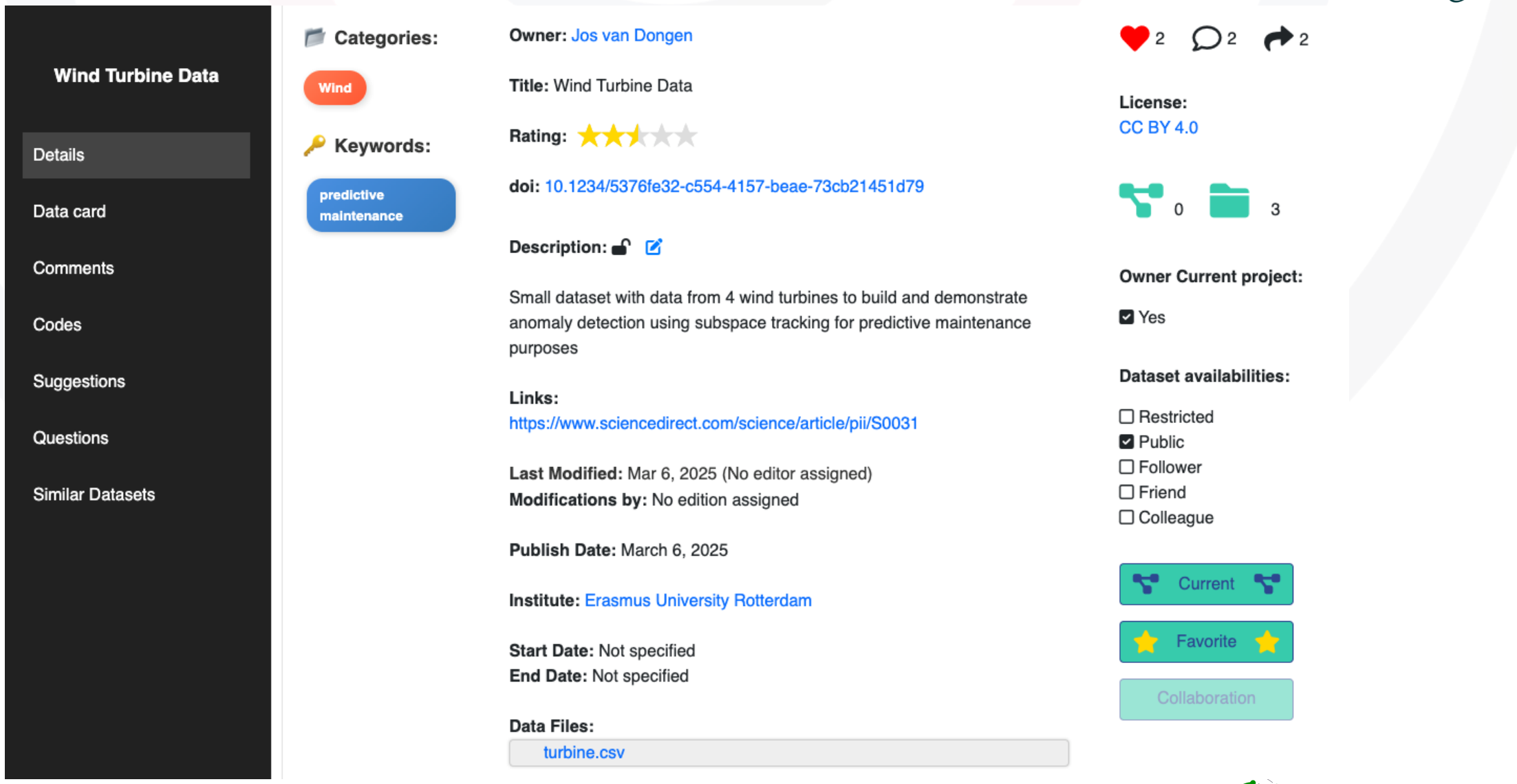
Simple UI: Datasets & Search







Full user control over everything



Restrict Dataset access

Erasmus University Rotterdam

Categories:

Thermal (Heat)

Keywords:

No keywords available.

Owner: ECDA ECDA

Title: Weather Dataset

Rating:

doi: 10.1234/9e1d575a-a026-47e6-869b-c1601067a2f6

Description:

Weather dataset in Parquet format

Links:
No links available.

Last Modified: Today at 1:46 PM (No editor assigned)
Modifications by: No edition assigned

Publish Date: March 15, 2025

0 0 0

License:
CC BY 4.0

0 3

Owner Current project:

Yes

Dataset availabilities:

- Restricted
- Public
- Follower
- Friend
- Colleague

User A

s available.

odified: Today at 1:46 PM (No editor assigned)
ations by: No edition assigned

Date: March 15, 2025

Institute: Erasmus University Rotterdam

Start Date: Not specified
End Date: Not specified

Views: 2 **Downloads:** 0

Data Files: Restricted access
To access of the dataset files, you should follow or be as a friend or colleague of ECDA ECDA.

Dataset availabilities:

- Restricted
- Public
- Follower
- Friend
- Colleague

Current

Favorite

Collaboration

Report Dataset

User B

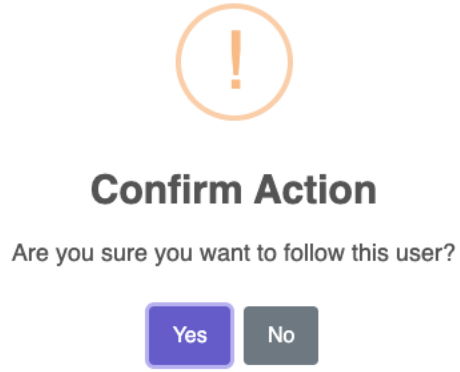
Erasmus Data Collaboratory

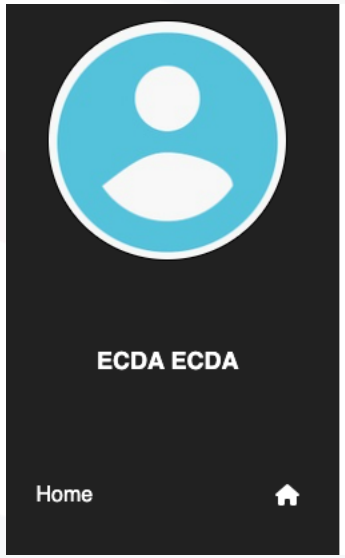
Dataset Access Levels

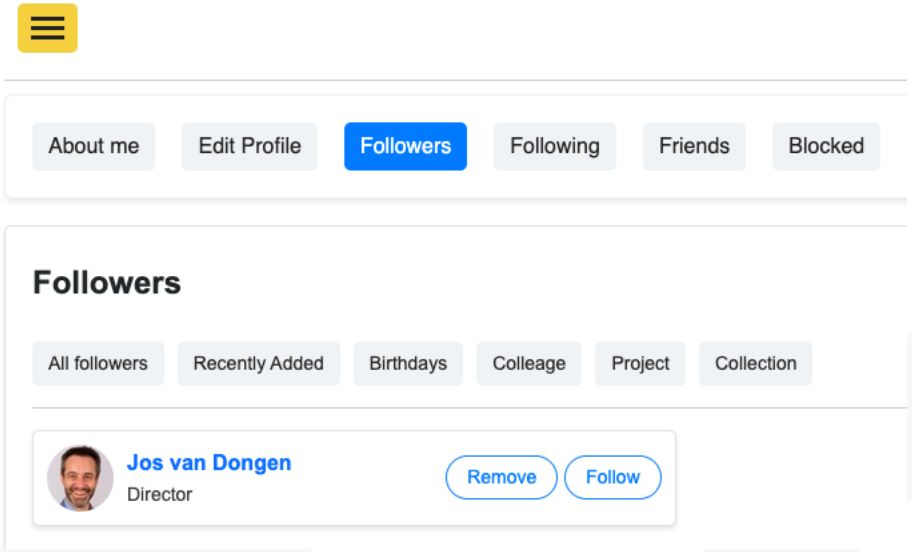
- Restricted: only you
- Public: anyone
- Follower: people who follow you
- Friend: only friends
- Colleague: project members

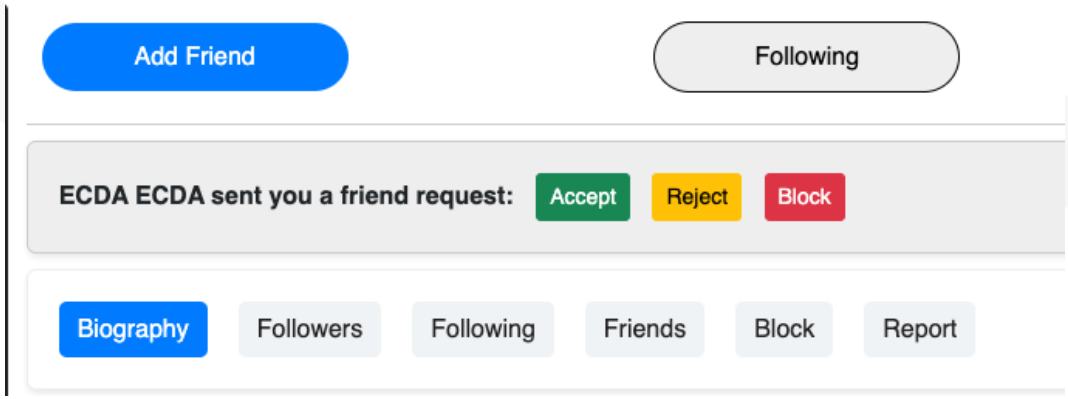
-> You can also block people

Following Users, Friend Requests



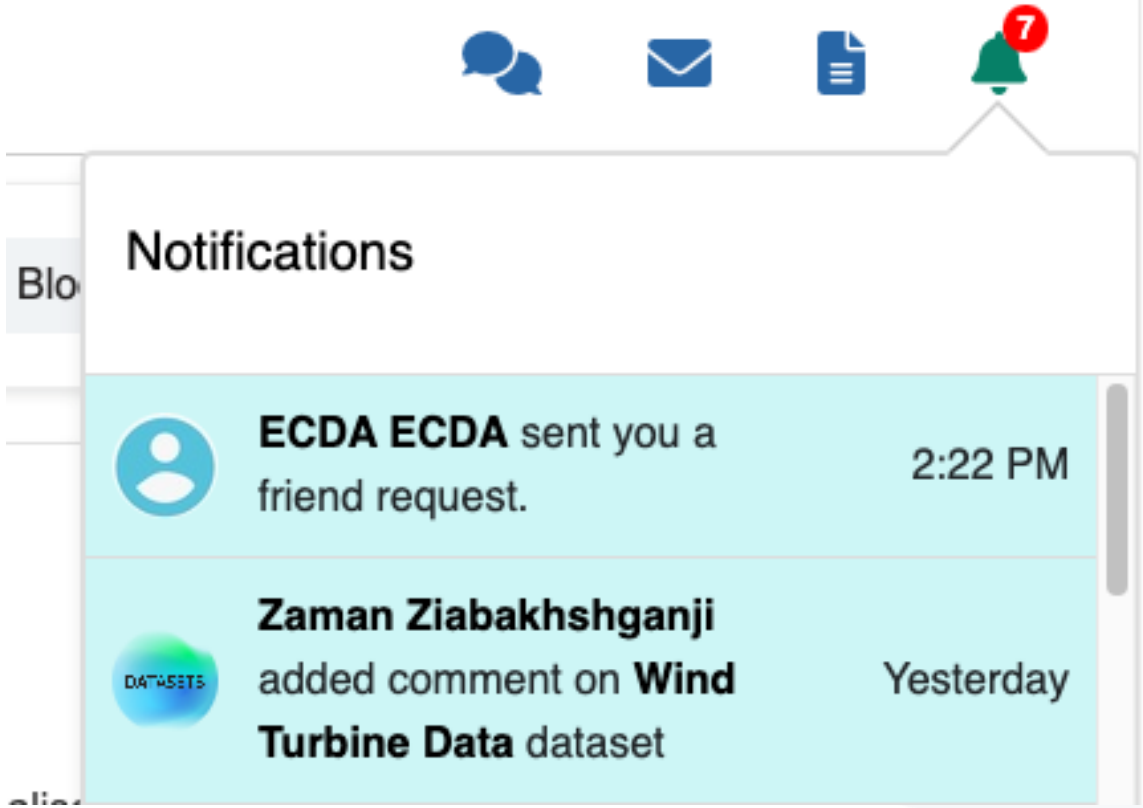


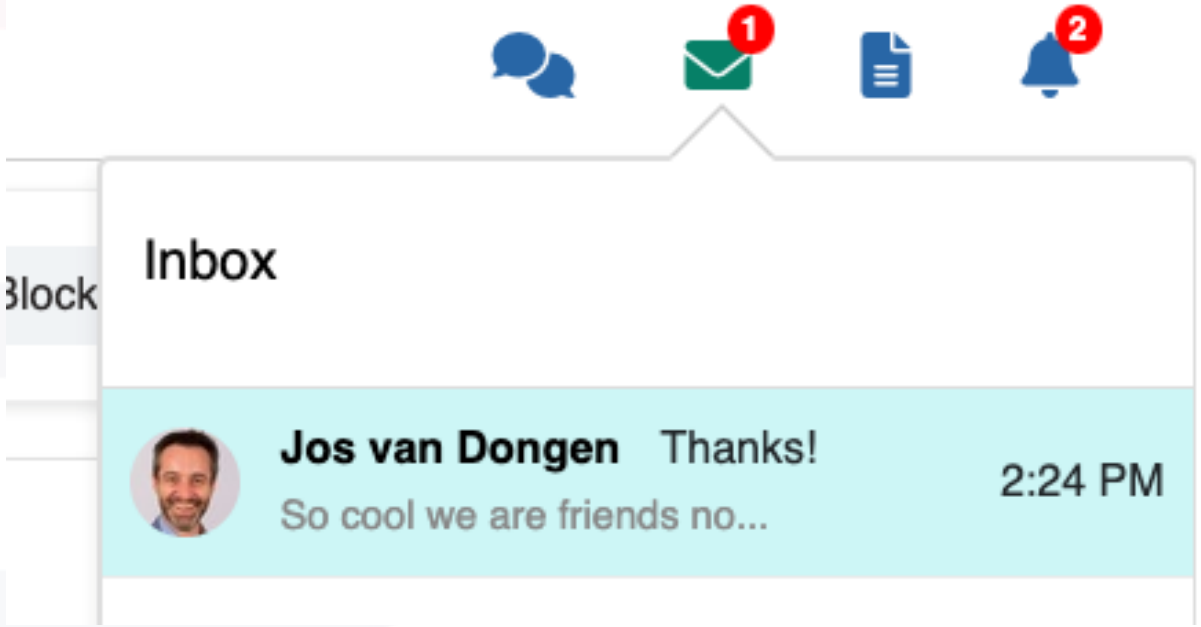




Integrated Mail/Messaging/Notifications

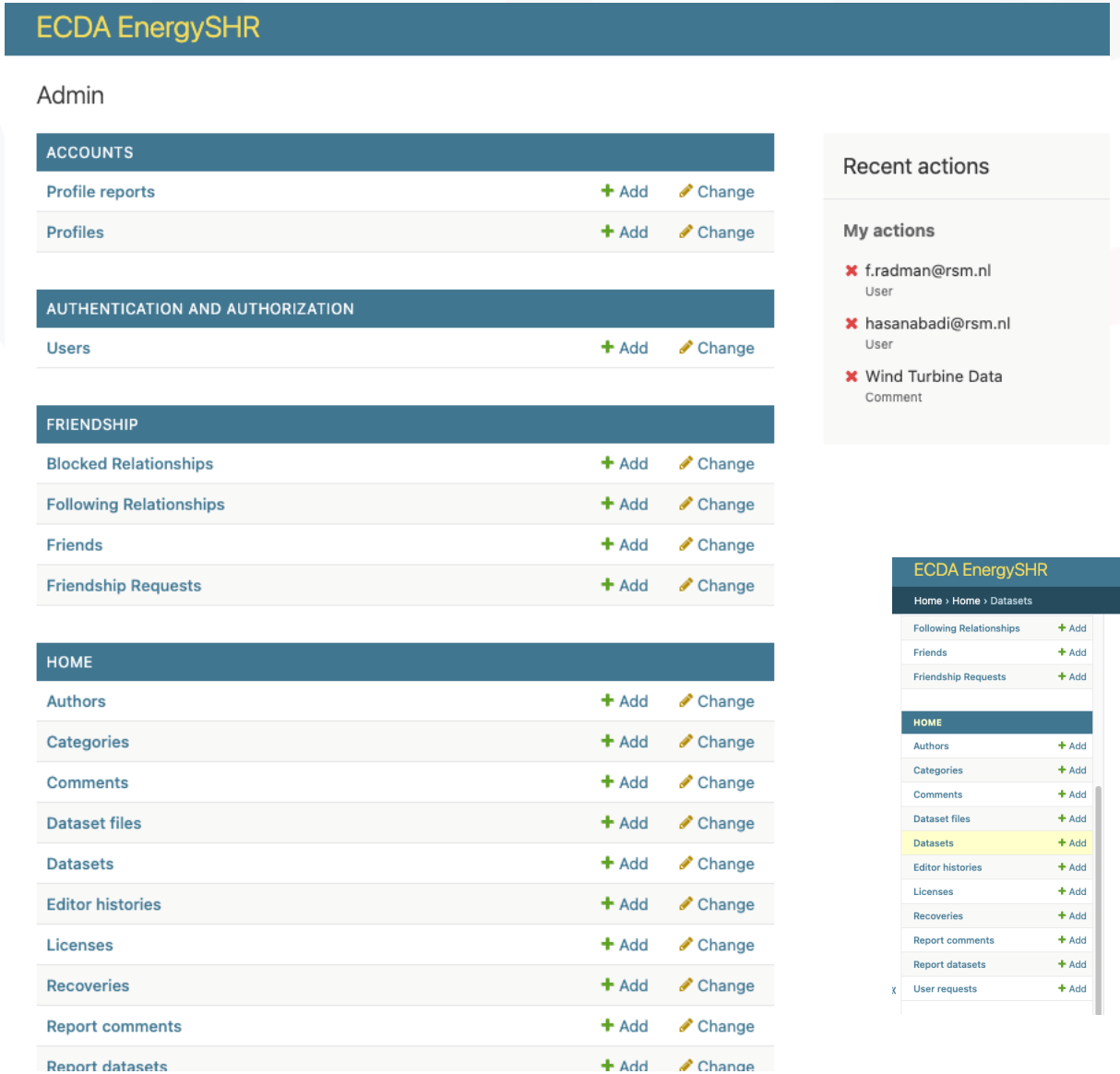
Erasmus University Rotterdam

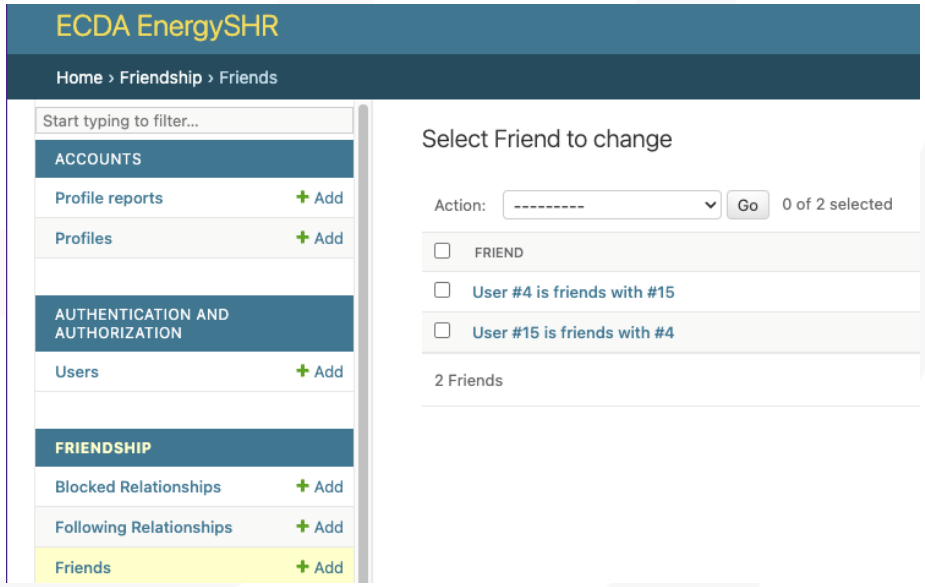


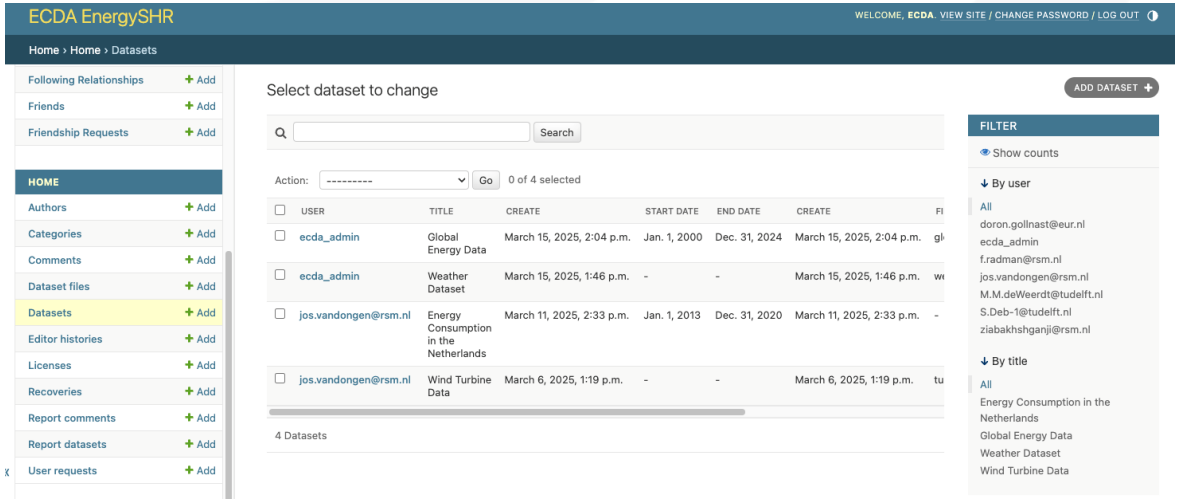


Administration backend

Erasmus University Rotterdam







Erasmus Data Collaboratory

Security: Surf Research Access Management

Erasmus University Rotterdam

SURF Research Access Management

Easy and secure access to research applications for research collaborations

Log in

New to SURF Research Access Management? Log in and create a collaboration with your institution.

SURF Research Access Management

Home > EnergySHR

EnergySHR

EnergySHR

Delft University of Technology

34 members & no groups

Member since July 31, 2023

Home > EnergyShare

Erasmus Data Collaboratory

EnergyShare

Short name: **energyshare**

Connected to 2 collaborations & available for 2 organisations

Open the application or Visit the website

Details & settings | Admins | Groups | Connected collaborations | Access rules | Connection requests

Security: Network Access Policies

Erasmus University Rotterdam

leafcloud | erasmus-data-collaboratory | admin@e

- Project
  - API Access
  - Compute
  - Rating
  - Volumes
  - Container Infra
  - Network
    - Network Topology
    - Networks
    - Routers
    - Security Groups
    - Load Balancers
    - Floating IPs
  - Orchestration
  - DNS

Project / Network / Security Groups / Manage Security Group Rul…

Manage Security Group Rules: EnergySHR (2cf79202-dbf9-4597-b861-da8cb9f8caaf)

+ Add Rule | Delete

Displaying 10 items

Direction	Ether Type	IP Protocol	Port Range	Remote IP Prefix	Remote Security Group	Description	Actions
Egress	IPv4	Any	Any	0.0.0.0/0	-	-	Delete
Egress	IPv6	Any	Any	::/0	-	-	Delete
Ingress	IPv4	TCP	22 (SSH)	[redacted]	-	-	Delete
Ingress	IPv4	TCP	80 (HTTP)		-	-	Delete
Ingress	IPv4	TCP	443 (HTTPS)		-	TUD HTTPS 3 (Private BClass2)	Delete
Ingress	IPv4	TCP	443 (HTTPS)		-	TUD HTTPS 1 (BClass)	Delete

Erasmus Data Collaboratory

Roadmap

Erasmus University Rotterdam

- Now – April 10: ironing out final issues
- April 10: Launch at Erasmus Data Summit
- Q2 2025: Packaging & deployment
- Q3 2025: Open Source
- Q4 2025: Analytics Integration (Jupyter? Superset?)

Questions?