



Erasmus Data Sandbox

Key Words: Shared Research Platform | Platform engineering | AI experimentation | Cross-Disciplinary Data Innovation

As data becomes increasingly central to research, education, and innovation, there is a growing need for secure, powerful, and flexible environments where users can explore ideas, test hypotheses, and build solutions. The Erasmus Data Sandbox was developed precisely to meet this need. The Sandbox is a core component of the Erasmus Data Collaboratory | House of AI (EDC), as it was one of the initial projects. Designed as a secure experimental environment, it offers researchers, commercial partners, and startups a space to experiment with data science and AI technologies. Whether building prototypes, exploring data, or testing models, users can treat the Sandbox as a "playground" to push boundaries and develop innovative approaches to real-world problems.

The Sandbox integrates cutting-edge data infrastructure, tooling, and collaboration features. EDC's platform engineer acts as the main developer and maintainer of the Sandbox. Through building the infrastructure, granting the correct access to the Sandbox storage, and allocating the appropriate computational power both from internal and external sources, the Sandbox provides access to computation power, datasets, algorithms, reusable code, and documentation within a shared environment. Data catalogue, analytics and engineering tools, AgenticAI platforms and Large Language Models are among such projects. These projects have already demonstrated the Sandbox's value in enabling rapid development and testing.

Alongside his engineering expertise, EDC's platform engineer also provides strategic consulting on computational power provisioning and platform suitability in the entire life cycle of data/AI-driven projects.

The Erasmus Data Sandbox showcases the forward-thinking approach the Erasmus Data Collaboratory | House of AI has. The framework and structure of the Sandbox further show the in-house knowledge and technical expertise EDC holds, leading towards AI and data-driven innovation. By providing expertise and tools, the Sandbox also helps create a collaborative knowledge base that can drive cross-disciplinary discovery into the future.

Stackholders: EDC

Tools/Tech used: Apache Hop, Apache Airflow, n8n, Apache Superset, Kubernetes, Proxmox, pfSense, Ceph, CrateDB

Sandbox Tech

