

Architectural Design of BERD Information Portal

Ahmed Saleh¹[\[https://orcid.org/0000-0003-0581-4655\]](https://orcid.org/0000-0003-0581-4655), and Klaus
Tochtermann^{1,2}[\[https://orcid.org/0000-0003-2471-2697\]](https://orcid.org/0000-0003-2471-2697)

¹ZBW – Leibniz Information Centre for Economics, Kiel/Hamburg, Germany

²Kiel University, Germany

Abstract: The management of Business, Economic, and Related Data can be a complex task involving various types of data and diverse research needs. BERD@NFDI develops a powerful platform to collect, process, analyze, and preserve such data in one place. The technical architecture of the BERD platform consists of several components, each responsible for specific functionalities that work together to deliver content and services to users. At a high level, BERD infrastructure is designed to align with the recommendations of the EOSC Interoperability Framework¹, which includes several layers beyond the technical layer, such as organizational interoperability, legal interoperability, and semantic interoperability. However, this document focuses solely on the technical interoperability of the BERD architecture.

At the base layer of the BERD infrastructure, there is a physical infrastructure that includes servers, storage devices, and networking equipment, providing computing power and storage space. On top of this, there is a software infrastructure that includes the operating system, web server, and application server, providing the necessary software components to run the platform. The data layer includes the database server and any other data storage systems used by the platform to store and manage data, such as user account information and data marketplace records. The presentation layer is responsible for rendering web pages and providing an intuitive and engaging user experience. The infrastructure that hosts the BERD platform is provided by a commercial cloud provider, and it consists of four main servers: the test server, the production server, the services server, and the mockups server. The BERD Platform is built on top of InvenioRDM², an open-source research data management platform. The services used by the platform include OpenSearch³, PostgreSQL⁴, Redis⁵, and RabbitMQ⁶. OpenSearch provides a distributed, multitenant-capable full-text search engine. PostgreSQL provides a database management system (RDBMS). Redis handles the cache of the BERD Platform, for example, storing user sessions and caching rendered pages. RabbitMQ holds the tasks for Invenio workers to execute. Invenio workers are background processes or threads that execute various tasks in the Invenio software stack, such as processing and indexing metadata records, handling user

¹<https://eosc-portal.eu/eosc-interoperability-framework>

²<https://inveniordm.web.cern.ch>

³<https://opensearch.org>

⁴<https://www.postgresql.org>

⁵<https://redis.com>

⁶<https://www.rabbitmq.com>

requests, and performing maintenance tasks. All of these services, with the exception of PostgreSQL, are deployed on the servers using Gitlab-CI and docker containers. Postgres is provided as a service by the cloud provider.

The architecture of the platform consists of four main components: crawlers, a data model, an ingestion service, and a machine learning component. Crawlers are responsible for harvesting data from online sources such as digital libraries, repositories, social media platforms, and other online repositories. The data collected by crawlers is then processed by a data model that aligns the data format and cleans it up as needed. The ingestion service is responsible for inputting the processed data into a database, where it can be queried and analyzed. Finally, a machine learning component can be used to enrich the data with additional insights and patterns.

Overall, the technical architecture of the BERD platform is designed to be scalable, reliable, and secure, while also providing a high-performance and user-friendly experience for visitors.

Keywords: BERD Infrastructure, Crawlers, Search Engine