Reengineering of CRIS at University of Novi Sad

Keywords

TeslaRIS, ElasticSearch, MinIO, Kibana, MongoDB

Introduction

Founded in 1960, the University of Novi Sad, serves as an autonomous hub for education, science, and arts in the province of Vojvodina. While its educational software infrastructure is robust, the research domain lacked adequate software support, leading to the initiation of the DOSIRD UNS project in 2009 (http://dosird.uns.ac.rs/). This initiative aims to develop a comprehensive software infrastructure tailored for research. Key activities encompass analyzing management requirements, researching European standards, creating research information systems, and integrating digital libraries of Ph.D. dissertations. The project emphasizes promoting open science activities, enhancing discoverability of the university's scientific outputs, fostering global research collaborations, preserving the university's scientific heritage, and elevating its academic rating. Ultimately, DOSIRD UNS aspires to meet local institutional mandates while aligning with international research standards, facilitating efficient research management, and amplifying the university's global academic influence. The main product of the DOSIRD UNS project is the CRIS UNS platform which was put into operation in 2010. More than 40 thousand research domain entities have been cataloged in the last 13 years. The platform evolved by including a lot of new features in this period. Piled up technical debt, development of web frameworks, and changes in scholarly communication required us to consider reengineering of the CRIS UNS platform. Reengineering the platform should offer numerous benefits, including cost optimization by reducing maintenance expenses and enhancing developer efficiency. The refactored code ensures faster system responses, facilitating easier feature enhancements. Additionally, modernized software adapts better to evolving scholarly communication trends and academic needs. The new platform hopefully might be used even out of University of Novi Sad, therefore we called that TeslaRIS. CRIS UNS will be just one customization of the TeslaRIS open source software.

The TeslaRIS Architecture

TeslaRIS is centered around a modular and scalable architecture implemented using the Spring Boot framework and follows the principles of a modular monolith (a.k.a. “modulith”). The system employs a PostgreSQL database as the primary storage for structured data. Elasticsearch is utilized for efficient indexing and searching capabilities, enhancing the speed and accuracy of data retrieval. Document storage is handled by MinIO, a high-performance object storage

solution. This ensures a scalable and reliable storage infrastructure for various types of documents associated with research information. Kibana is employed as the dashboard and visualization tool, providing administrator users with insightful data representations and facilitating data-driven decision-making, as well as aiding in the development and debugging process. The use of aforementioned established technologies ensures a robust and scalable infrastructure, fostering interoperability and enhancing the overall research information ecosystem. The data flow within the system is orchestrated to provide maximum consistency and efficiency. Harvested data from external systems is initially stored in MongoDB, processed, and subsequently transferred to the primary PostgreSQL database. Conversely, data export follows a reverse path, ensuring synchronization and data integrity. The system is structured into several interconnected modules, each serving a distinct purpose in the research information lifecycle.

- **Core Module** forms the backbone of the CRIS, providing essential functionalities and services. It encapsulates the core business logic, CRUD operations for basic entities, configuration, and data models necessary for managing and processing research information.
- **Import Module** is responsible for efficiently collecting data from external sources such as other CRIS systems, global publication platforms, files, etc. It will implement the ETL (extraction, transformation, loading) process. The extraction submodule will be extensible with support for harvesting information from different sources (REST APIs, OAI-PMH endpoints, file formats, SKG-IF, etc). The result of this submodule will be temporarily stored in a MongoDB intermediate database. If it is necessary, system administrators can consolidate and clean up data, before its transformation and loading into the TeslaRIS database. This asynchronous harvesting process ensures seamless integration with external systems and enhances data interoperability.
- **Export Module** is tasked with disseminating research information to external systems or exporting in numerous formats for the users' needs. The module can accept the search query or exact list of records which should be exported, as well as needed exporting configuration (for instance, file format, protocol type, record formats, etc.). In the case of a permanent interoperability with other platforms established by using periodic queries, it periodically fetches data from the primary PostgreSQL database, transforming it, and publishing it to the MongoDB intermediate storage. This design choice facilitates an efficient and standardized data exchange mechanism with external entities.
- **Evaluation Module** is responsible for collecting numerous indicators and assigning to the research results or results' sources. Moreover, this module should enable utilizing the process of evaluation, from the definition/configuration of evaluation criteria, over application for the evaluation, to the automatic, semi-automatic or manual assessment of researchers and their work. Although,
- **Report Module** facilitates internal and external reporting of research activity of organizational units at multiple levels of the organizational hierarchy, as well as at the level of an individual (researcher) or project. The module will enable definition of reports by providing configuration of initial report data (defined by a query or by selection of records), formats and layouts. This module will also enable reporting about usage

statistics of the platform for the needs of maintainers. The platform will log messages and ELK stack (ElasticSearch, Logstash, Kibana) will be used for visualization.

- **PhD Dissertations Library Module** is tightly integrated with the core module, offering specialized functionalities for handling doctoral theses and associated processes within the CRIS. The module will support the process of application for defending a PhD dissertation, and process of promotion of a PhD. Moreover, it should enable linking other research outcomes to a PhD dissertation.

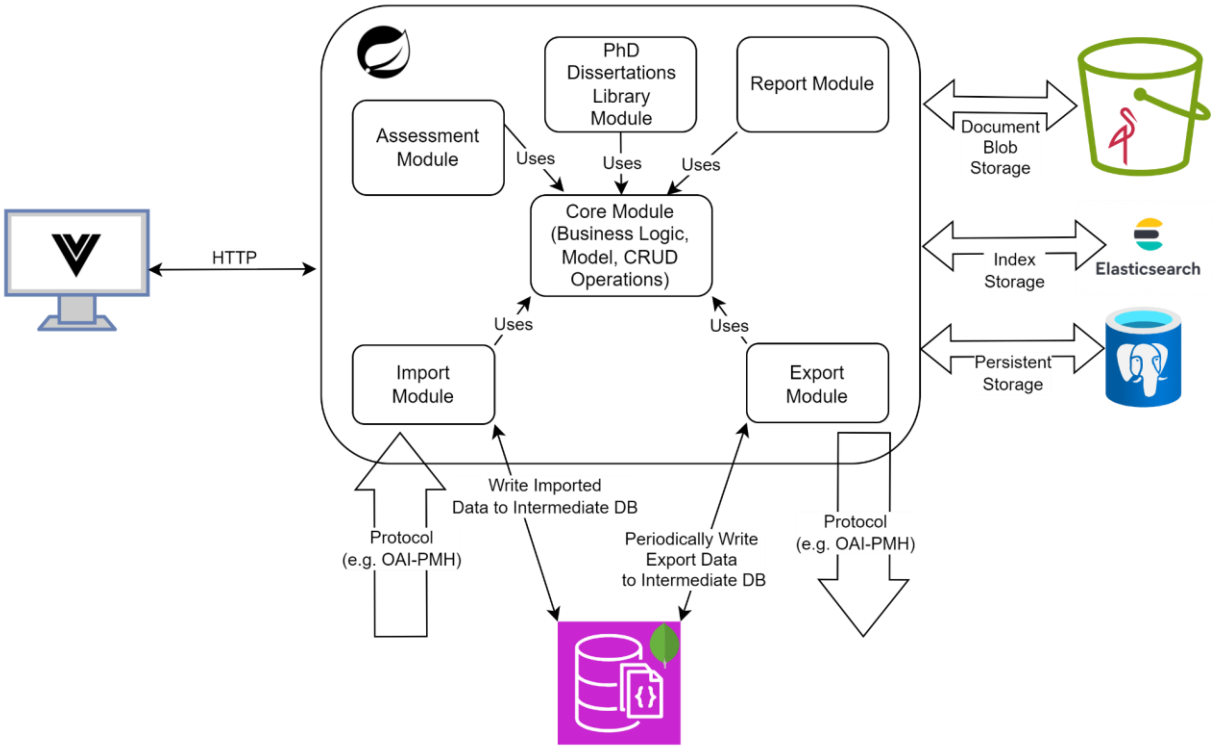

Figure 1 - TeslaRIS's Architecture Diagram

Data model

The TeslaRIS data model is aligned with basic ideas of the refactored CERIF model (CERIF core and scholarly publication module). The data model expressed in the Plant UML notation can be found at https://github.com/sci2zero/TeslaRIS-backend/blob/feature/docs/docs/model.puml.

Implementation

The implementation will be done in phases:

- The first phase includes the implementation of the Core, Import and Export modules. The Import module will be used for migration of over 40.000 records from the previous platform.

- The second phase will include the implementation of Assessment and Report modules
- The last phase will include integration of the PhD digital library within the CRIS system.

The TeslaRIS application will be published as an open source software. The source code is available at https://github.com/sci2zero/TeslaRIS-backend and https://github.com/sci2zero/TeslaRIS-frontend.

The first release should be published later this year. A demo of the ongoing implementation will be presented at CRIS 2024, hopefully followed by useful feedback from the community. We are expecting that the first phase of the implementation will be ready for demo at the conference.

Acknowledgement

This research has been supported by the Ministry of Science, Technological Development and Innovation (Contract No. 451-03-65/2024-03/200156) and the Faculty of Technical Sciences, University of Novi Sad through project “Scientific and Artistic Research Work of Researchers in Teaching and Associate Positions at the Faculty of Technical Sciences, University of Novi Sad” (No. 01-3394/1).