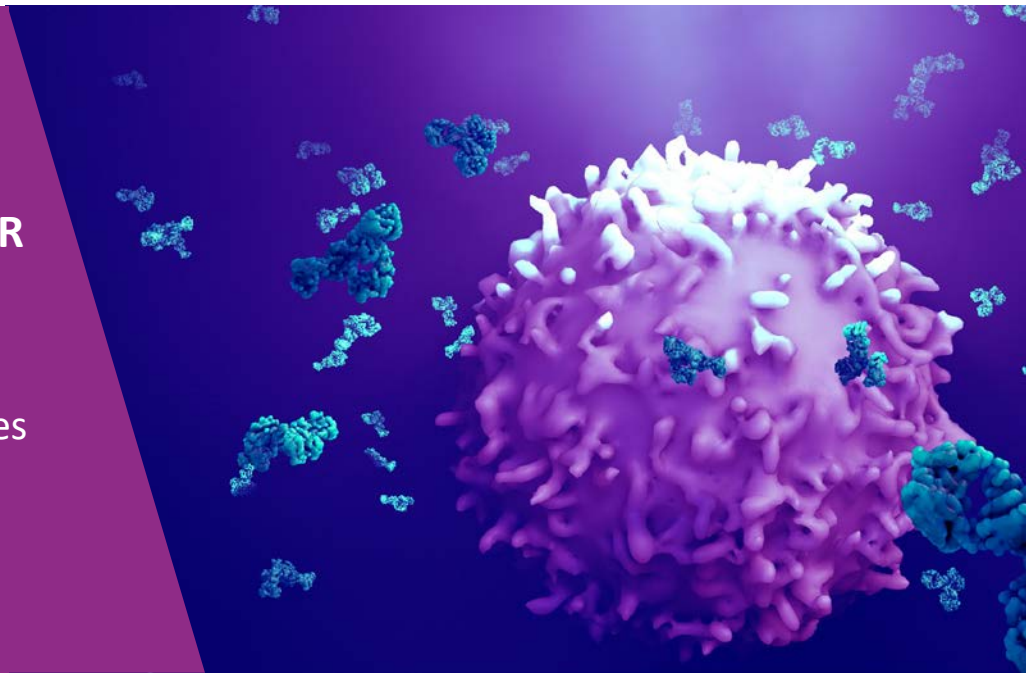
imSAVAR

IMMUNE SAFETY AVATAR

Nonclinical mimicking of the immune system effects of immunomodulatory therapies



Deliverable 5.6
Catalogue of Datasets

DELIVERABLE REPORT

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 853988.
The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and JDRF INTERNATIONAL.

innovative medicines initiative

efpia

JDRF IMPROVING LIVES. CURING TYPE 1 DIABETES.

Abstract

This document reports on the implementation of a dataset catalogue for imSAVAR using the ELIXIR-LU Data Catalog system. In particular, it focuses on the technology that has been adopted to provide an end-to-end system for users to access and find data. The end-to-end solution includes the ELIXIR Data Catalog, DAISY (Data Information System), LFT (LCSB File Transfer), and REMS (Resource Entitlement Management System). We have populated the catalogue with two datasets for which we have received the data. As we receive more datasets, they will be added to the catalogue.

Document Information

Deliverable Report	**D5.6: Catalogue of Datasets**
Date	**26.02.2024**
Report prepared by	**University of Luxembourg** Muhammad SHOAIB, Venkata Pardhasaradhi SATAGOPAM **Information Technology for Translational Medicine** Andreas Kremer, Nils Christian Verena Kopfnagel (MHH)
Project	**imSAVAR - Immune Safety Avatar: nonclinical mimicking of the immune system effects of immunomodulatory therapies** Grant Agreement No.: 853988 (IMI2-2018-15-04)
Project Coordinator	**Fraunhofer-Gesellschaft zur Foerderung der angewandten Forschung e.V.** Prof. Dr. Dr. Ulrike Köhl **Novartis Pharma AG** Dr. Jonathan Moggs
Type	**Deliverable Report | Public**

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 853988. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and JDRF INTERNATIONAL.

Table of Contents

1. Methods

1.1 Metadata Collection Process

Collecting metadata about datasets was one of the primary tasks towards the development of the data catalogue. We undertook the following steps to gather metadata about the datasets:

- **Meta-Data Collection Preparation:** As a crucial first step in creating the imSAVAR dataset catalogue, we initiated the process by collecting metadata for the datasets. This was essential to provide users with a comprehensive overview of the available data. We employed a structured approach of using REDCap to this process.
- **Instrument Development:** To streamline the metadata collection, we leveraged the Research Electronic Data Capture (REDCap) platform[1], a robust tool for managing research data. Within REDCap, we meticulously designed and configured an instrument specifically tailored for collecting metadata.
- **Instrument Creation and Testing:** Once the REDCap instrument was created, it went through a rigorous testing phase to ensure its functionality and usability. Our team, consisting of data scientists and project managers, collaborated to refine the instrument for efficiency and accuracy.
- **Testing Workflow:** During this phase, we examined every aspect of the instrument, including data entry forms, validation rules, and user interface design. We conducted extensive testing to identify and rectify any potential issues.
- **User Training:** As data managers, we drafted comprehensive documentation for the data collection team.
- **Requesting Data via REDCap:** With the fully functional REDCap instrument in place, we initiated requests to data providers. This step was fundamental to collect the actual data and establish a standardized process for data submission. UNILU coordinated with data providers to ensure a clear understanding of the data submission process. We provided guidelines and instructions to streamline the submission of metadata via the REDCap instrument.
- **Data Validation and Quality Assurance:** As data manager, we implemented data validation checks within the instrument to enhance data quality. This included checks for consistency, completeness, and data format, ensuring that the metadata received were of high quality.

In conclusion, the metadata collection process was meticulously planned and executed by our team, aligning with our expertise in data science, project management, and data stewardship. By leveraging REDCap for metadata collection, we ensured the catalogue's success in providing high-quality and standardized metadata for stakeholders within our organization. This process laid a solid foundation for the catalogue's development and ongoing maintenance.

1.2 Metadata Conversion from REDCap to DATS

We converted metadata collected via REDCap into DATS (Data Tag Suite) standard used by ELIXIR-LU Data Catalog. Our team, consisting of data scientists and project managers, worked collaboratively to ensure the accurate and complete representation of the data within the DATS framework. The conversion process was executed meticulously to maintain data quality and consistency. Once the metadata was successfully converted, it became more findable and accessible, aligning with our organization's goal of efficient data management and sharing.

[1] https://imsavar-redcap.lcsb.uni.lu/

2. Results

We have successfully completed the implementation of a comprehensive data catalogue for the imSAVAR project. This data catalogue serves as a central repository for all project-related data, providing a structured and organized approach to data management. The end-to-end solution, which integrates various components such as DAISY, Data Catalog, LFT, and REMS, represents a robust framework that ensures compliance with the General Data Protection Regulation (GDPR). By implementing this solution, we are able to efficiently manage data accessibility and streamline access management processes.

2.1 Components of the End-to-End System

2.1.1 DAISY - GDPR Compliance Tool

DAISY serves as the cornerstone for GDPR compliance within the imSAVAR project. It is a versatile tool that records and manages processing activities related to datasets. This includes crucial details such as data types, sensitivity classes, use restrictions, and a comprehensive log of accesses, providing an unparalleled level of transparency. Functioning as the ground truth for the system, DAISY ensures that the data management processes align with GDPR regulations.

2.1.2 LFT – IBM Aspera-Based Sharing Tool

LFT (LSCB File Transfer) is an integral part of the end-to-end system, serving as the IBM Aspera-based sharing tool. Its robust infrastructure is designed to facilitate secure and high-speed data transfer, ensuring that data sharing within the imSAVAR project is efficient and complies with security standards.

2.1.3 Data Catalog - Advertisement Portal

Data Catalog[2] plays a pivotal role as the central advertisement portal, acting as the primary entry point for users seeking access to datasets. It interfaces seamlessly with DAISY, extracting detailed metadata to display to users. Keycloak integration allows users to log in using various identity providers, such as UniLu or LCSB accounts, ORCID, or LifeScience login. Users can explore datasets and request access, initiating the streamlined workflow.

[2] https://datacatalog.elixir-luxembourg.org/r/project/imsavar

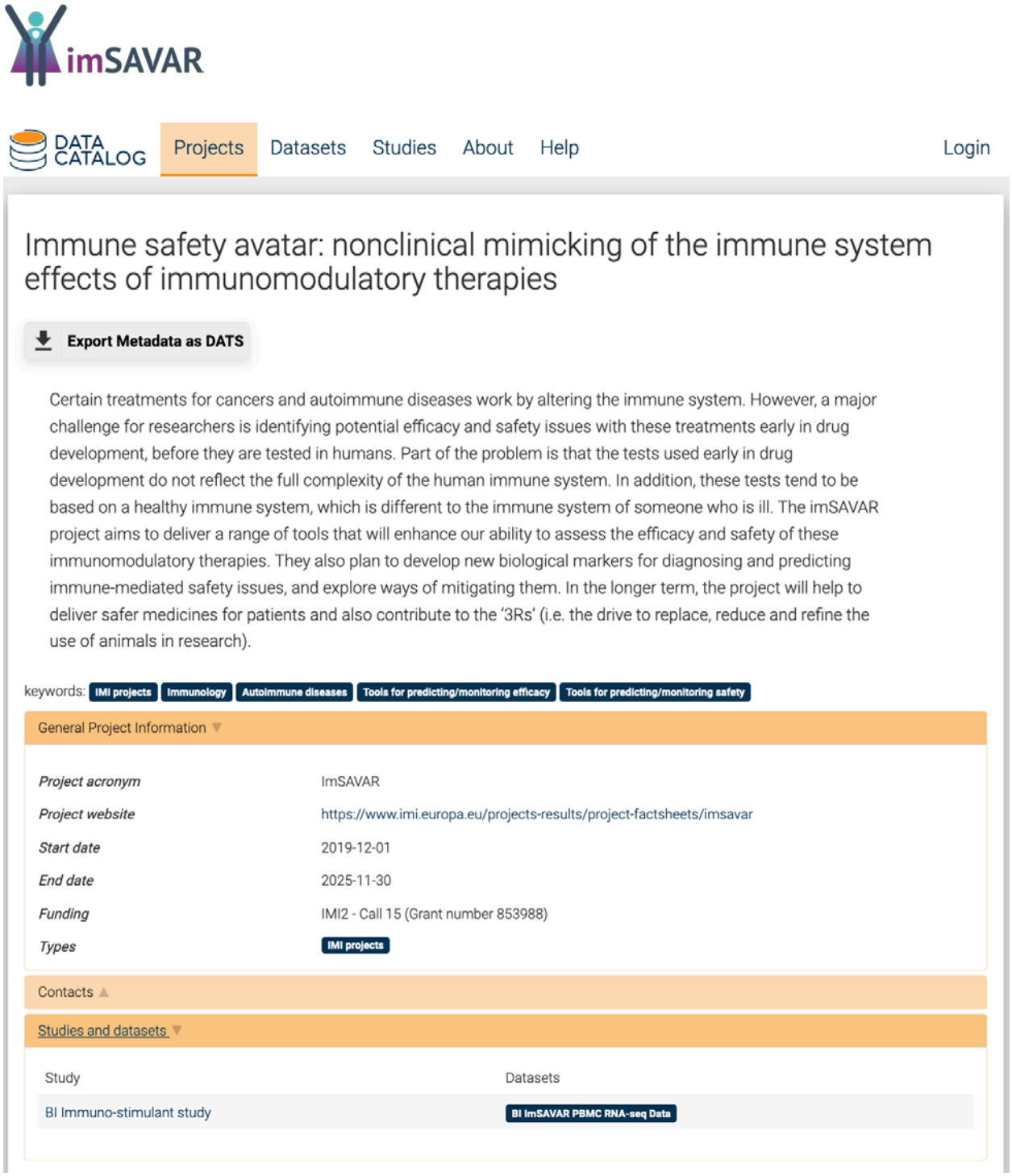


Figure 1: Screenshot from Elixir Data Catalogue

2.1.4 REMS - Access Management System

REMS is the access management powerhouse, facilitating the intricate process of approving or disapproving access requests. It provides a collaborative platform for Data Access Committee (DAC) members to exchange opinions and make informed decisions. Every step within REMS is meticulously recorded, contributing to an exhaustive audit log that ensures GDPR compliance.

2.2 End-to-End Workflow

2.2.1 User Access Request

- Users log in to Data Catalog using Keycloak and various identity providers.
- The portal provides a user-friendly interface for users to explore datasets and initiate access requests.

2.2.2 Access Approval Process (REMS)

- REMS facilitates real-time communication among DAC members, fostering a collaborative decision-making environment.
- Comprehensive audit logs are maintained within REMS to document every aspect of the access approval process.

2.2.3 User Notification and Data Download

- Users receive timely notifications via email upon the approval of access requests.
- Seamless integration with Data Catalog ensures that users can log in and download approved data. DAISY and LFT collaborate to verify user access and generate time-restricted password-protected links for data download.

2.2.4 Audit Log and GDPR Compliance

- **REMS Audit Log:** Every step within REMS is recorded, creating a detailed audit log. This ensures transparency, accountability, and GDPR compliance.
- **DAISY Repository:** DAISY serves as the centralized repository for GDPR-related information, providing a comprehensive view of processing activities, access details, and data restrictions.

2.3 Future Expansion and Scalability

The implemented system is designed with scalability in mind, anticipating the addition of more datasets as the project evolves. The flexible architecture ensures that the system can seamlessly accommodate the diverse research needs of imSAVAR. Two datasets have been successfully FAIRified as part of the imSAVAR project. The FAIRification process ensures that the dataset aligns with the FAIR principles, making it more discoverable and accessible for researchers within the imSAVAR community. Data catalogue ensures the data findability and accessibility however, we still need to make datasets interoperable and reusable, therefore, we are currently in the process of FAIRifying another dataset that has been received for processing. This involves enhancing the dataset's metadata, ensuring proper documentation, and making it compliant with FAIR standards. The ongoing FAIRification efforts contribute to the overall goal of creating a robust and FAIR-aligned dataset repository.

The comprehensive implementation of the data catalogue for the imSAVAR project signifies a significant advancement in data management capabilities. The end-to-end solution not only ensures GDPR compliance but also streamlines the data accessibility and sharing process. As imSAVAR continues to grow, the adaptability and scalability of this integrated system will play a pivotal role in supporting the dynamic research objectives of the project. Furthermore, metadata about the datasets that cannot be stored on central location because of data restrictions.

3. Conclusion

The integrated use of DAISY, LFT, Data Catalog, and REMS in our data management processes ensures a robust and comprehensive framework that is fully compliant with General Data Protection Regulation (GDPR). This system not only enhances transparency in data processing activities but also provides a secure and streamlined approach to data access, promoting responsible and compliant data management practices. By leveraging the capabilities of these technologies, we are able to effectively manage data across various systems and ensure data integrity throughout the entire data lifecycle. Additionally, the implementation of this framework allows for efficient data analysis and reporting, enabling informed

decision-making and improved business outcomes. Overall, our data management approach, supported by the integrated use of DAISY, LFT, Data Catalog, and REMS, empowers organizations to confidently navigate the complexities of data governance and compliance in today's digital landscape.

Acknowledgement

The authors would like to express their gratitude to the Innovative Medicines Initiative 2 Joint Undertaking (JU) for the financial support of this research under grant agreement No 853988. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and JDRF INTERNATIONAL.