

# Data Management for Systems Medicine: The SMART-CARE Joint Environment

Friedemann RINGWALD<sup>a,1</sup>, Felix CZERNILOFSKY<sup>b</sup>, Aleksei DUDCHENKO<sup>a</sup>,  
Matthias GANZINGER<sup>a</sup>, Sascha DIETRICH<sup>b</sup> and Petra KNAUP<sup>a</sup>

<sup>a</sup>*Institute of Medical Biometry and Informatics, Heidelberg University Hospital*

<sup>b</sup>*Department of Hematology, Heidelberg University Hospital*

**Abstract.** For a research project on mass spectrometry, a streamlined, harmonized and robust analytical pipeline is built to predict tumor recurrence. By means of standardization all steps from sample collection, analysis, proteome, and metabolome analysis are harmonized. Challenges like non-central identifiers and distributed data are overcome with a centralized high-performant IT-platform in combination with a pseudonymization service and harmonization.

**Keywords.** systems medicine, research platform, mass spectrometry

## 1. Introduction

SMART-CARE (Systems Medicine Approach to Stratification of Cancer Recurrence) is a combined research project consisting of several clinical and biomedical partners in Heidelberg, Germany. Its goal is to build a robust mass spectrometry (MS)-based pipeline standardizing all steps from sample collection and analyte extraction to data processing by MS in standard operating procedures to ensure reproducible proteome and metabolome analysis. In a next step, the pipeline is deployed in a systems medicine approach to predict tumor recurrence. The various clinical partners and the aim of central data collection require a complex sample management assuring correct and unambiguous sample identification. Therefore, data management is a central point in this joint research project.

## 2. Methods

A high-performant, stable and constantly accessible joint platform for clinical, laboratory and analytical data has to be created to establish data management. Further, providing an identity management solution and pseudonymization service for handling samples and clinical data is an important step to ensure correct allocation of samples. Thus, a central pseudonym for each patient must be provided. In addition, harmonization of clinical and laboratory data is important but challenging due to the participation of multiple partners from different hospitals and laboratories with individual data conventions. To ensure

---

<sup>1</sup>Corresponding Author, Friedemann Ringwald; E-mail: [friedemann.ringwald@med.uni-heidelberg.de](mailto:friedemann.ringwald@med.uni-heidelberg.de)

comparability of data across different institutions, harmonization is necessary with the requirement of flexible specifications. A core data set is defined in a first step. Finally, data protection requirements including follow-up data for patients must be fulfilled to stay within legal boundaries.

### 3. Results

As underlying IT platform, OpenBis [1], a flexible open-source framework for managing and analyzing complex data is used to handle all data provided from clinics, laboratories and analytical partners. The software provides form-based solutions for sample management and data repository. The clinical partners each have different spaces where they can access, create, and edit data of their corresponding patients. Based on valuable input of piloting clinical partners, forms and masks are defined. Through a web-service interface, the identity management software Mainzelliste [2] is integrated into the openBis forms. While Mainzelliste has been used in similar approaches [3], to our knowledge it has not been so tightly integrated with openBis. Our approach assures consistent identification of samples and clinical data since data privacy requirements prohibit use of identifying data (patient name, birthdate, e.g.). All project members including hospitals, laboratories and data analysis teams work on this central openBis platform and share their data and insights, setting a very interdisciplinary and novel environment. Furthermore, the entire system is hosted on a dedicated high-performance server to store and modify large data sets reliably. Additionally, GPU support is included to provide a runtime environment for training and applying models from data analysis. Challenging are the individual data conventions of the data provided by partners from different clinics. A harmonization over all collectable data is not possible since different diseases are considered. Therefore, where applicable, openBis forms are configured to reflect disease-specific harmonized data definitions.

### 4. Discussion

Although openBIS requires some effort in training end-users, we made good overall experiences with the software so far. In an iterative process, the configuration of openBIS is improved and extended to the remaining partners. In the next step, laboratory and analysis partners will be integrated into the process to attain an analytical pipeline that fits the purpose of predicting relapse of malign diseases.

### References

- [1] Bauch A, Adamczyk I, Buczek P, Elmer F-J, Enimanev K, Glyzowski P, Kohler M, Pylak T, Quandt A, Ramakrishnan C, Beisel C, Malmström L, Aebersold R, Rinn B. openBIS: a flexible framework for managing and analyzing complex data in biology research. *BMC Bioinformatics*.2011;12, 468.
- [2] Lablans M, Borg A, Ückert F. A RESTful interface to pseudonymization services in modern web applications. *BMC Med Inform Decis Mak*.2015;15, 2.
- [3] Scheible R, Rusch S, Guzman D, Mahlaoui N, Ehl S, Kindle G. The NEW ESID online database network, *Bioinformatics*. 2019;35, 24.