



The use of the EUDAT repository to store clinical trials in a secure and compliant way

The EUDAT repository and services will be used for the secure, GCP (Good Clinical Practice) compliant and transparent storage of clinical trials data. For such a safe and accessible storage of clinical trials, an authentication service (AAS) manages the access rights for users; a PID service refers users, digital objects and associated documents to each other. In addition, the linking to metadata and the data type registry is necessary. Leading European clinical researchers are centred in the European Clinical Research Infrastructures Network (ECRIN), whose members and other researchers will access the repository to analyse data of different European trials.

The Scientific Challenge

Randomized clinical trials (RCT) are an important step to bring treatments from preclinical development to the patient. But many scientific and technical challenges exist for clinical trials, like the need for innovation in trial design and for more objective interpretations of trial outcome data. There is a gap in the translation of basic scientific discoveries into clinical trials and of clinical trials into medical practice. Although, biomedical sciences provide an unprecedented supply of information for improving human health, clinical trials data do not participate in the activities of the research environment in an important way. After the conclusion of a clinical trial, most raw data is withdrawn and inaccessibly archived and only statistical summary results are published. A repository for clinical trial data (raw data in anonymised form) would become the first step for the provision of this data to the research community for analysis. The course of clinical trials is determined by a detailed study protocol; patient data is collected by many investigators at different sites using electronic Case Report Forms (eCRF). Increasingly data from biobanks, nutritional and genetic data and data from electronic health records (EHR) are involved and exist in different formats (see Fig. 1 below).

After the end of the study, data is analysed using statistical software and study results may be published. Nonetheless, the clinical trials raw data is stored in isolated archives without metadata enrichment and without links and references to preclinical data, trial documents, publications, and analysis.

Who benefits and how?

Clinical trials are the gold standard for research with human data; they deliver data of highest quality and scientific relevance. To open this data source for research communities is of great benefit for scientific research in general. The result of this pilot activity will increase transparency by linking trials data to reports and publications and make trial data accessible for analysis by researchers, providing a means to compare them with data of other trials or pre-clinical experimental results.

The main beneficiary is the clinical research community who can improve trials planning and conduct. Other beneficiaries are the research community in general, and all researchers who want to compare their data to the results of clinical trials.

Technical Implementation

The aim of the pilot is to evaluate EUDAT services for the storage and sharing of data from clinical trials. The focus is on B2SAFE as a service for archiving of clinical trials data. In contrast to other pilots, which are using B2SAFE as a simple backup and replication solution for an existing community repository (e.g. DSPACE, Fedora), we use B2SAFE for long-term storage by directly importing data into B2SAFE. Because access to the data is only

granted to the community data manager, the requirement for restricted access is ensured.

B2SAFE is already installed and hosted at the Forschungszentrum Juelich (FZ Juelich). At the Heinrich-Heine University Duesseldorf (HHU), an iRODS instance was installed on a HHU server and then configured (Fig. 1). B2SAFE data is replicated to other storage sites, like, for example the Karlsruhe Institute of Technology (KIT), to ensure the creation of copies of the B2SAFE content although this replication and synchronisation is done with non-encrypted data.

The import of datasets into B2SAFE is done by synchronisation via iRODS-to-iRODS between the HHU server at our community location and the one at FZ Juelich with the B2SAFE instance. From FZ Juelich B2SAFE is replicated with other B2SAFE locations. This synchronisation is done via iRODS rule engine.

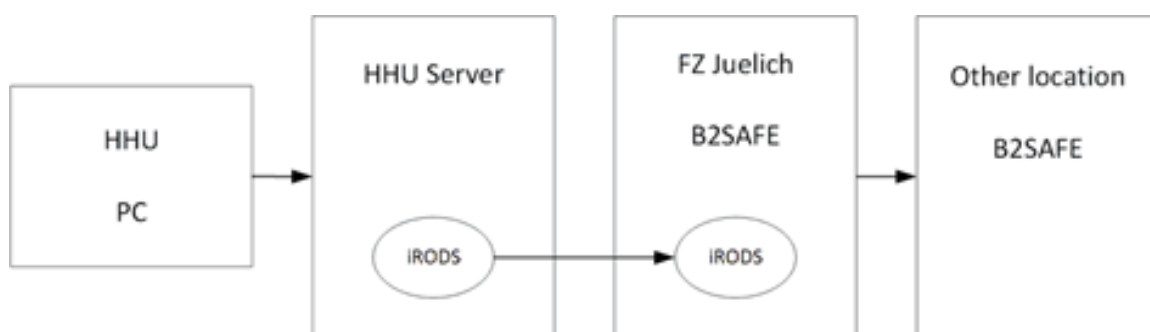


Fig. 1: Installation of iRODS server at HHU (Heinrich-Heine University)

B2SHARE is a repository for shareable digital objects. We use B2SHARE to distribute some anonymised datasets and publications that are open, or link to these documents. The pseudonymised datasets and the study documents are archived under restricted access conditions in B2SAFE, whereas a subset of the data is anonymised and shifted into B2SHARE for publication and sharing (Fig. 2). For the B2SHARE service no installation at our location was necessary; we are using the general EUDAT service.

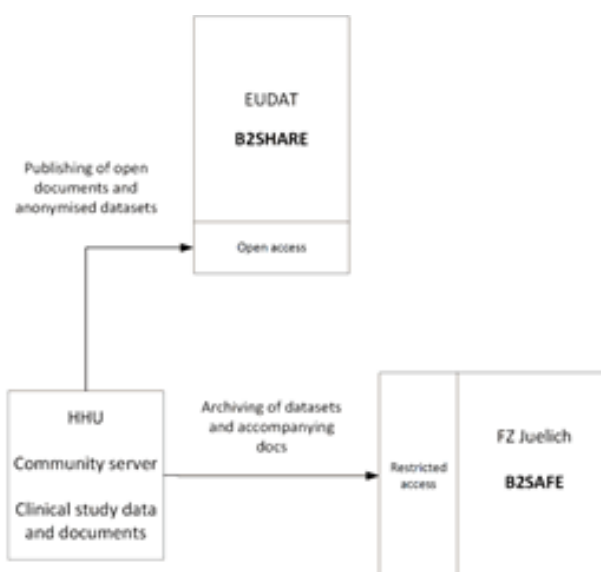


Fig. 2: Concept for the storage of protected data in B2SAFE with access restriction, and for open data in B2SHARE for open access

Preliminary Results

In summary, the first impression of the pilot is that archiving of datasets and document can be achieved simply and without problems. The proof-of-concept for storing clinical trials data in a safe manner was positive, if one uses encrypted data. The iRODS / B2SAFE implementation worked well and is robust. The use of B2SAFE for protected data is cumbersome and requires additional steps, like the encryption of data at the community site. The ingest step is done without a data management interface and uses only iRODS, which is intricate to use. Currently all access and transfer of data has to go through the local community server and encryption must be conducted at the local community site.

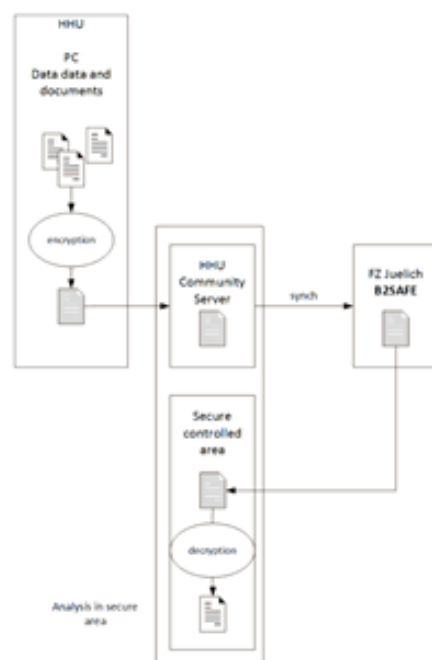


Fig. 3: Concept for the encryption of study data, so that data transfer, B2SAFE synchronisation and data analysis can take place in a protected manner. (synch=synchronisation)

Contacts

- Wolfgang Kuchinke, HHU, Kuchinke(at)med.uni-duesseldorf.de
- Sander Apweiler, Research Centre Jülich, sa.apweiler(at)fz-juelich.de

Further Information

[Scientific Challenges behind the Pilot](#)

[To learn more about the EUDAT/ECRIN collaboration click here](#)

[Read more](#)