Management probenassoziierter Daten im onkologischen Kontext der Medizinischen Hochschule Hannover

Management of bio sample related data in the oncological context of the Hannover Medical School

Markus KERSTING [a1], Svetlana GERBEL [a], Jana PROKEIN [a], Hans LASER [a], Jörg HAIER [b], Norman SCHOENFELD [a], Thomas ILLIG [c]
[a] *Center for Information Management, Hannover Medical School*
[b] *Comprehensive Cancer Center Hannover, Hannover Medical School*
[c] *Hannover Unified Biobank, Hannover Medical School*

Abstract. The Hannover Medical School (MHH) has established the central Hannover Unified Biobank (HUB) and a central Enterprise Clinical Research Data Warehouse (ECRDW). These central research services, the clinical context and all corresponding data flows are now to be tighly integrated into all oncological use cases, especially the upcoming Comprehensive Cancer Center (CCC).

Keywords. Biobank, linical Research Data Warehouse, Enterprise IT Infrastructures, Secondary Use of Medical Data, Oncology, Bio samples, Data

Background

The Hannover Medical School (MHH) has established the Hannover Unified Biobank (HUB) [1], one of the largest university biobanks in Germany. In cooperation with the Institute of Pathology, approximately 1.2 million tissue samples and 960,000 liquid samples are stored and managed in the HUB.

The Enterprise Clinical Research Data Warehouse (ECRDW) of the MHH enables the secondary use of medical data and, with over 3 billion data points, operates as a central interdisciplinary platform for research-relevant usage scenarios [2].

In order to achieve a top position in oncology as a Comprehensive Cancer Center (CCC) [3], it must be possible to call up differentiated and complex query options and statistics on biosamples of oncologically treated patients currently available at the MHH, including clinical, molecular, therapy-associated and experimental data. To achieve this objective, a demand-oriented provision of all necessary data on oncological biosamples for use in research and CCC should be established.

Methods

An inventory of the existing central clinical documentation systems of the MHH, which contain oncological data, as well as an analysis of the associated processes was carried out. In detail, interfaces of the biobank information and management system

[1] Corresponding Author

(CentraXX)[2] and the ECRDW, as well as possible options for data provision processes were evaluated. In order to use research-relevant data in use cases, such as the CCC in this case, data integration routes (using ETL[3]) were created or existing routes were adapted.

To provide structurally and semantically aligned data, systems and interfaces based on international standards were used. In concrete terms, this project relied on the communication infrastructure based on HL7 (Ensemble[4]) established at the MHH in the clinical field. Data from the clinical routine is transmitted to the communication server via HL7 interfaces and then integrated into the ECRDW via ETL for research-relevant use cases. Systems without primary relevance to clinical routine (e.g. tumor registers, biobanks) are addressed directly via the ECRDW technology and integrated into the data repository of the ECRDW. Data modeling as well as data validation and cleansing (e.g. by plausibility checks and harmonization by conversion of data types) are also performed using ECRDW technology.

Results

The necessary requirements were clarified in several consultations and workshops between the interdisciplinary domain experts (e.g. tumour centre, oncological departments, biobank) and IT experts of the MHH Center for Information Management.

The HUB delivers sample data according to an established HUB basic data set to the ECRDW. This consists of donor and sample-related data, including information on quality and processes. Within the framework of the CCC, an adapted version of this basic data set was developed for a harmonized presentation of sample data from pathology, HUB and the CCC project partner, the UMG Biobank in Göttingen[5]. Thus, the following harmonized information can be transmitted to the ECRDW for all samples from pathology and HUB and thus be used in the context of the CCC:

- Unique patient ID (PID) for the confusion-free linkage with clinical primary data of the MHH
- Time of sample collection
- Material type and processing coded according to SPREC [4] including time stamps for complete documentation of processing times and sample quality
- Sample quantity and availability
- Storage location (e.g. biobank/pathology)

The final, seamless integration of the new ONKOSTAR[6] oncological documentation system into the system is almost complete.
The established central infrastructure for the CCC and the associated data flows are illustrated in *Figure 1*.

[2] https://www.kairos.de/produkte/centraxx-bio/

[3] Extract-Transform-Load

[4] https://www.intersystems.com/de/produkte/ensemble/

[5] http://www.biobank.med.uni-goettingen.de/

[6] https://www.onkostar.de/de/index.php

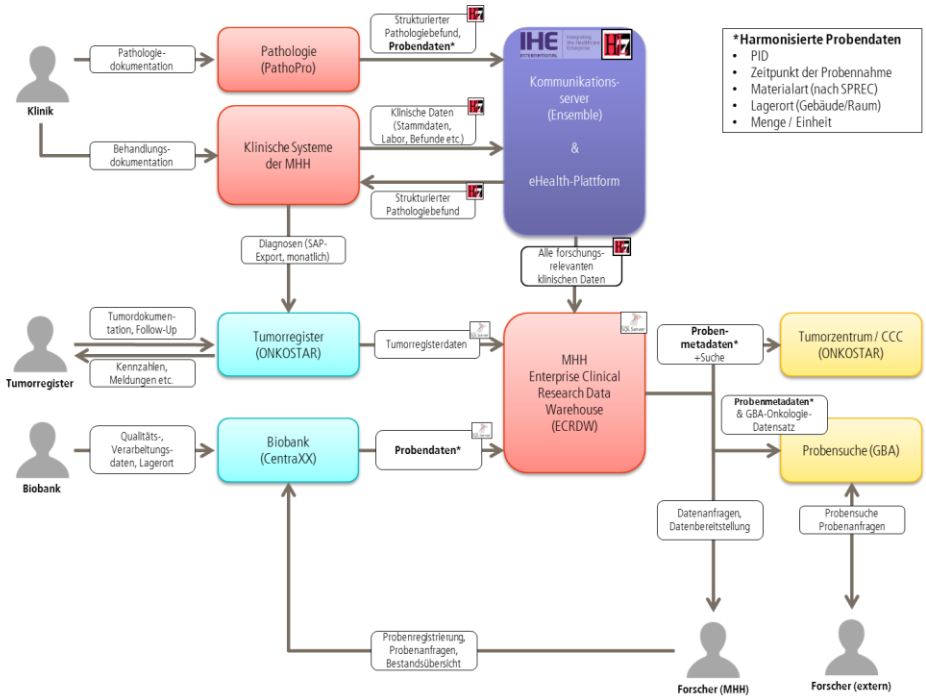

Figure 1. Central infrastructure and the associated data flows in the oncological context

Discussion

The execution of queries and the generation of statistics on the samples available at the MHH from patients undergoing oncological treatment is already available by linking the central data integration and analysis platforms Biobank and ECRDW [5]. The seamless integration of a central documentation system for tumor documentation will enable a future CCC to perform precisely tailored queries and statistics as well as to harmonize processes in order to meet all requirements and requirements associated with a CCC.

In order to enable data exchange across facilities, the connection via IHE infrastructure is planned.

References

1 Bernemann I, Kersting M, Prokein J, Hummel M, Klopp N, Illig T. Centralized biobanks - a basis for medical research. Bundesgesundheitsblatt, Springer 2016 Mar;59(3):336-43. doi:10.1007/s00103-015-2295-2

2 Gerbel S, Laser H, Schönfeld N, Rassmann T. The Hannover Medical School Enterprise Clinical Research Data Warehouse: 5 Years of Experience: 13th International Conference, DILS 2018, Hannover, Germany, November 20-21, 2018, Proceedings. Data Integration in the Life Sciences, 01/2019: p. 182-194. doi:10.1007/978-3-030-06016-9_17

3 Comprehensive Cancer Center. http://www.ccc-netzwerk.de/; Letzter Zugang: 2019-07-23

4 Betsou F, Lehmann S, Ashton G, Barnes M, Benson EE, Coppola D, DeSouza Y, Eliason J, Glazer B, Guadagni F, Harding K, Horsfall DJ, Kleeberger C, Nanni U, Prasad A, Shea K, Skubitz A, Somiari S, Gunter E. Standard preanalytical coding for biospecimens: defining the sample PREanalytical code. Cancer Epidemiol Biomarkers Prev. 2010 Apr;19(4):1004-11. Epub 2010 Mar 23. doi: 10.1158/1055-9965.EPI-09-1268

5 Gerbel S, Kersting M, Laser H, Bernemann I, Illig T. How can health care data and biobanks be adapted and linked for research – Establishing an Enterprise Clinical Research Data Warehouse. GMDS 2018. 63. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. (GMDS). Osnabrück, 02.-06.09.2018; 2018. doi:10.3205/18gmds131