Innovative use of Information for Clinical Care and Biomarker Research

*Stefan Schulz[1], Markus Kreuzthaler[1], Berthold Huppertz[2], Karine Sargsyan[2], Werner Leodolter[3], Markus Pedevilla[3], Dagmar Becker[4], Richard Ackbar[5], Robert Fasching[5], Thomas Pieber[5,6]; [1]**Medical University of Graz** – Institute for Medical Informatics, Statistics and Documentation, [2]**Medical University of Graz** – Biobank Graz, [3]**KAGes** – Steiermärkische Krankenanstaltengesellschaft m.b.H., [4]**SAP SE**, [5]**CBmed GmbH** – Center for Biomarker Research in Medicine, [6]**Medical University of Graz** - Division of Endocrinology and Metabolism*

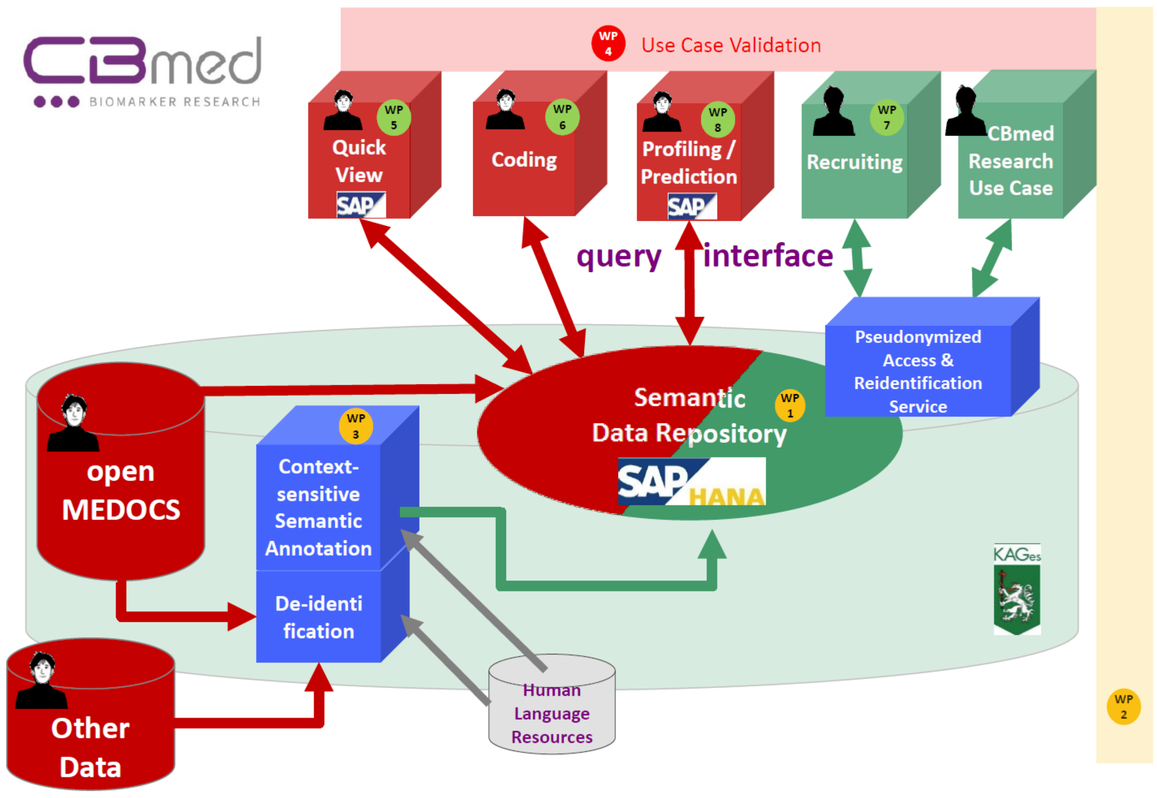
Context

Electronic records in health care institutions constitute an increasingly interesting data treasure for primary and secondary use scenarios**. Innovative use of Information for Clinical Care and Biomarker Research –** Project 1.2 in CBmed, investigates methods and develops technologies for the management of clinical and research data. It pursues the goal to enable innovative re-use of electronic records with structured and unstructured data from a broad range of heterogeneous sources, leveraging current technology in the fields of **data semantics**, **biomedical terminology**, **natural language processing** [1], **big data management** and **predictive content analytics**.

Basic technologies

A **semantic data repository** is developed in WP1. It pools data from heterogeneous sources and is embedded in SAP-HANA [2]. Besides data in their original format, it contains rich semantic extracts from free-text clinical narratives, using standardized vocabularies (e.g. SNOMED CT) and a complex data model able to represent clinical data in context.

Architecture for Clinical and Research Use Cases



The extraction process is addressed by WP 3 (Semantic technologies and resources). Using text mining and language technologies, supported by domain and site specific vocabularies it fills the semantic data repository with structured content. This repository thus constitutes a rich source of clinical information that can be queried by a broad range of use cases. WP4 provides tools and resources for qualitative and quantitative assessment to guarantee quality of the extraction and retrieval results.

Use cases

We distinguish clinical from research use cases. The architecture addresses different data protection requirements (see figure). Legal and ethical issues regarding data ownership and protection are addressed by WP2, as well as an enhancement of existing named entity de-identification [3] services for structured and unstructured data, meeting international standards (e.g. HIPAA).

Research use cases focus on the building of patient cohorts according to semantic criteria. E.g., WP 7 "Recruiting" allows for context-based delineation of (de-identified) patient cohorts for research. Suitable patients for a planned study are selected, according to phenotypic and genotypic information. A broad range of applications to support biomarker-related research is possible. Other use cases are formulated by the biobank which is interested in using clinical data for the identification of suitable biosamples.

Clinical use cases support clinicians and administrators in their daily work: WP 5, "Patient Quick View" aims at providing a synopsis of the entire clinical history of a single patient, to support content filtering especially for cases of numerous treatment episodes such as in chronically ill patients. Another use case is the support of clinical coding (WP6), where terminology codes (e.g. ICD-10, MEL, ICPM) are automatically proposed to improve coding quality and to facilitate the identification of relevant features like clinical scores. Another use case (WP8) allows for patient profiling and prediction. It identifies specific patient profiles out of the wealth of available clinical data, based on statistical evaluation of comparable disease progressions.

[1] Friedman, C., & Elhadad, N. (2014). Natural language processing in health care and biomedicine. In Biomedical Informatics (pp. 255-284). Springer London.
[2] Färber, F., May, N., Lehner, W., Große, P., Müller, I., Rauhe, H., & Dees, J. (2012). The SAP HANA Database--An Architecture Overview. IEEE Data Eng. Bull., 35(1), 28-33.
[3] Neamatullah, Ishna, et al. (2008). Automated de-identification of free-text medical records. BMC medical informatics and decision making, 8(1), 32.

Acknowledgements: Work done in "CBmed" was funded by the Austrian Federal Government within the COMET K1 Centre Program, Land Steiermark, and Land Wien.

funded by:

CONTACT
CBmed GmbH
CENTER FOR BIOMARKER RESEARCH IN MEDICINE
Stiftingtalstrasse 5 8010 Graz
Phone: +43 316 385 28801
office@cbmed.at