

**CBmed GmbH**  
**Center for Biomarker Research in Medicine**  
**Programme: COMET – Competence Centers for Excellent Technologies**  
**Programme line: K1-Centers**  
**COMET sub-project 1.2, duration and type of project:**  
**Innovative Use of Information for Clinical Care and Biomarker Research (IICB), 01/2015 – 12/2018, multi-firm**

**Language technologies unlock clinical data for biomarker research**

In order to make routine clinical data available for biomarker research, structured content has to be extracted from the mostly text-based clinical information. The CBmed project 1.2 "Innovative Use of Information for Clinical Care and Biomarker Research" uses text mining and bespoke German terminology resources to analyse clinical jargon, feed semantic extracts into a data warehouse as a basis for clinical queries.

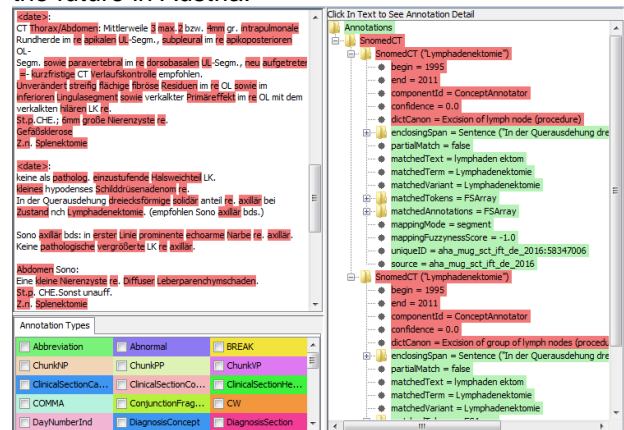
**Free-text documentation in electronic health records**

Each clinical encounter is mirrored by a variety of data stored in information systems. This includes current complaints, pre-existing illnesses, disease courses, diagnoses, findings as well as surgical interventions, medications and lifestyle factors. A vital part of this information is only available in the form of minimally structured discharge letters, problem lists and finding reports. These texts are characterised by highly compact terminology. Tailored to inter-expert communication between experts it presents a major challenge to computational analyses. Phrases like: „St. p. TE eines exulc. sek.knot.SSM li US dors. 5/11 Level IV 2,4 mm Tumordurchm. Sentinell LK ing. li. tumorr.“ are difficult to read even for non-specialist physicians. For biomarker research, however, it is essential to have relevant clinical data in a structured, standardised and therefore universally searchable form.

**Human language technologies and resources**

At CBmed, a team of medical informatics specialists are working on software solutions that extract, classify and transfer information from

clinical texts into structured data repositories. The prototypes developed in 2017 combine rule-based methods with machine learning techniques using commercial software and in-house developments. An important resource is a lexicon developed and maintained by the Medical University of Graz, which maps the local medical jargon to the international terminology standard SNOMED CT, which will also be introduced in the future in Austria.



**Fig 1: Analysis of a part of a clinical summary. Here, the word "lymphadenectomy" is recognized and annotated with the SNOMED CT term "excision of groups of lymph nodes (procedure)"**

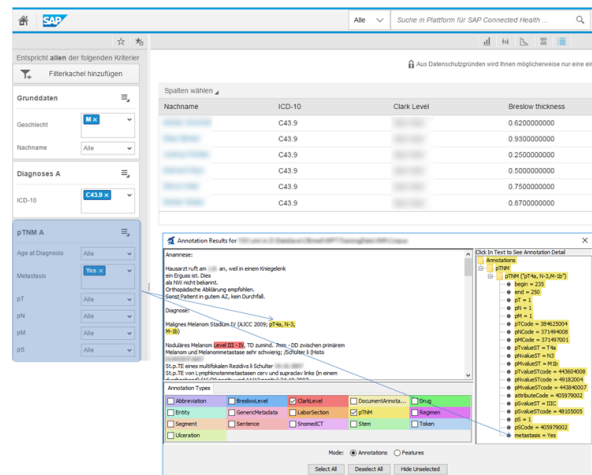


## Effects and benefits

The aim of this semantic processing of routine medical data is the creation of a "structured patient profile", following international standards and implemented as a "clinical data warehouse" in the SAP HANA system hosted by the Styrian healthcare provider KAGes. The system facilitates the recruitment of candidates for clinical trials. Equally, clinical search criteria such as tumour type, severity, comorbidities, therapies and time intervals can help find the right samples in a biobank. Another benefit of the semantic analysis of clinical texts is the improvement of clinical data quality.

Within less than two years, a complex and powerful platform that semantically enriches clinical routine data has been created and demonstrated on several prototypes. Use cases that extend beyond the application to biomarker research include decision support, data visualisation and quality assurance. Thus, a sustainable evolution of the system is endorsed by the

hospital operator KAGes, the Medical University of Graz and the software company SAP.



**Fig 2: Search interface of the Clinical Data Warehouse on SAP HANA. Here, extracted and automatically coded information can be retrieved from medical reports along with their temporal context.**

### Kontakt und Informationen

K1-Centre CBmed

CBmed GmbH  
Stiftingtalstrasse 5, 8010, Graz  
+43 316 385 28801  
[office@cbmed.at](mailto:office@cbmed.at)  
[www.cbmed.org](http://www.cbmed.org)

### Project coordinator

Univ.-Prof. Dr.med. Stefan Schulz

### Projektpartner

Organisation	Land
KAGes	Austria
SAP	Germany

**Weitere Informationen zu COMET – Competence Centers for Excellent Technologies:** [www.ffg.at/comet](http://www.ffg.at/comet)

Diese Success Story wurde von der Konsortialführung/der Zentrumsleitung zur Verfügung gestellt und zur Veröffentlichung auf der FFG-Website freigegeben. Für die Richtigkeit, Vollständigkeit und Aktualität der Inhalte übernimmt die FFG keine Haftung.