The Clinical Data Intelligence Project

A smart data initiative

Daniel Sonntag · Volker Tresp
Sonja Zillner · Alexander Cavallaro
Matthias Hammon · André Reis
Peter A. Fasching · Martin Sedlmayr
Thomas Ganslandt
Hans-Ulrich Prokosch · Klemens Budde
Danilo Schmidt · Carl Hinrichs
Thomas Wittenberg · Philipp Daumke
Patricia G. Oppelt

Introduction

Individuals respond differently to drugs and sometimes the effects are unpredictable. An underlying factor is that differences in the DNA contribute significantly to variation in the treatment responses of individuals. A general conclusion is that the intersection of genomics and medical clinical data has the potential to yield a new set of diagnostic tools and decision support methods that can be used to individualise and optimise patient therapy and care.

As a basis for the development of a new set of diagnostic and decision support tools, we concern ourselves with the topic of "data intelligence" which refers to the interaction with large amounts of data in rich, semantically meaningful ways, going beyond search possibilities to create the path from data to information to knowledge. The term *clinical data intelligence* in particular takes medical data from heterogeneous resources into account, combines the extracted information, and generates medical clinically-relevant knowledge about patients or treatments. The healthcare sector has been identified as one of the main areas to benefit from the recent trend towards data intelligence and large scale data analytics [16, 21].

In the KDI project funded by the Federal Ministry for Economic Affairs and Energy (BMWi), we assume that clinical big data analytics needs to focus on the clinical decision processes to become smart data (Fig. 1). With longitudinal data from many patients and with complete patient information, new reasoning processes based on big data analysis and data predictions can potentially be implemented. Many patients are needed to capture the whole complexity of the medical domain. Complete information about each individual patient is needed to minimise the effect of con-

Fig. 1 Towards smart data

DOI 10.1007/s00287-015-0913-x

Daniel Sonntag
German Research Center for Artificial Intelligence
Stuhlsatzenhausweg 3, 66123 Saarbrücken
E-Mail: daniel.sonntag@dfki.de

Volker Tresp · Sonja Zillner
Siemens, München

Alexander Cavallaro · Matthias Hammon · André Reis
Peter A. Fasching · Martin Sedlmayr · Thomas Ganslandt
Hans-Ulrich Prokosch
Uker, Erlangen

Klemens Budde · Danilo Schmidt · Carl Hinrichs
Charité, Berlin

Thomas Wittenberg
Fraunhofer IIS, Erlangen

Philipp Daumke
Averbis, Freiburg

Patricia G. Oppelt
IFG, Erlangen

Published online: 04 August 2015

Abstract

This article is about a new project that combines clinical data intelligence and smart data. It provides an introduction to the "Klinische Datenintelligenz" (KDI) project which is founded by the Federal Ministry for Economic Affairs and Energy (BMWi); we transfer research and development results (R&D) of the analysis of data which are generated in the clinical routine in specific medical domain. We present the project structure and goals, how patient care should be improved, and the joint efforts of data and knowledge engineering, information extraction (from textual and other unstructured data), statistical machine learning, decision support, and their integration into special use cases moving towards individualised medicine. In particular, we describe some details of our medical use cases and cooperation with two major German university hospitals.

founding factors[1]. These are obviously important areas of research; it is, however, crucial to be able to locate appropriate datasets for research in the first place and to use domain knowledge to integrate disparate, typically distributed datasets. The tasks of identifying potential impacts on the practice of personalised care and constructing authoritative knowledge bases for clinical decision support pose a lot of challenges in legal and technological terms. To overcome technological challenges of data access and data integration, we require knowledge engineering in the medical domain by using technical ontologies and metadata standards [51]. For example, truthfully representing patient information requires the use of adequate ontologies and terminologies which have been developed in recent years in medical knowledge engineering and in the context of the development of guidelines. In addition, the information on which the physicians base their decisions is often not contained in structural form but at best in various textual (and possibly hand-written) reports. Information extraction from the available raw source data is required to make this information available for a subsequent analysis. Due to the difficulties in extracting information from texts such as medical reports and other source data, image data and OMICS data[2] in particular can possibly provide additional important information for further processing. Finally, statistics and statistical machine learning methods provide a well established framework for the modelling and analysis of "actions" in a specific medical context. Here one can potentially build on many years of previous work on medical knowledge representations, i. e. applying large corpora of medical ontologies, taxonomies and dictionaries, e. g. SNOMED (Systematized Nomenclature of Medicine), ICD (International Classification of Diseases), RadLex (covering radiology-specific terms), and FMA (Foundational Model of Anatomy ontology) to be used beyond accounting purposes.

Due to the described complexity, the clinical data intelligence project requires joint efforts of knowledge engineering, information extraction from textual data, iconical data (pertaining to images), and other unstructured data as well as statistical machine learning approaches. We claim that "Clinical Data Intelligence" is a perfect field for exploiting the economic potential and future development in Germany, where knowledge engineering, information extraction, and statistical machine learning can benefit from one another. Potential applications of the Ministry's smart data initiative (Fig. 2) are as follows: first, the prediction of actions (e. g. diagnoses or procedures) to support a physician's decisions by modelling medical practice; second, an analysis of the benefits of medical actions in terms of a final outcome; and third, a system that provides the physician with indications (which of the potential actions under consideration to select as to generate the greatest patient benefit). We also discuss some practical aspects of the German research project KDI which involves two major German university hospitals, namely the Charité in Berlin and the University Hospital in Erlangen (Uker).

[1] Consider an automatic program attempting to assess the effectiveness of drug X, from population data in which drug usage was a patient's choice. Data show that gender differences influence a patient's choice of drug as well as their chances of recovery (Y). In this scenario, gender Z confounds the relation between X and Y since Z is a cause of both X and Y (Wikipedia).

[2] The English-language neologism OMICS refers to genomics, proteomics or metabolomics and aims at the collective characterisation and quantification of pools of biological molecules that translate into the structure, function, and dynamics of an organism.

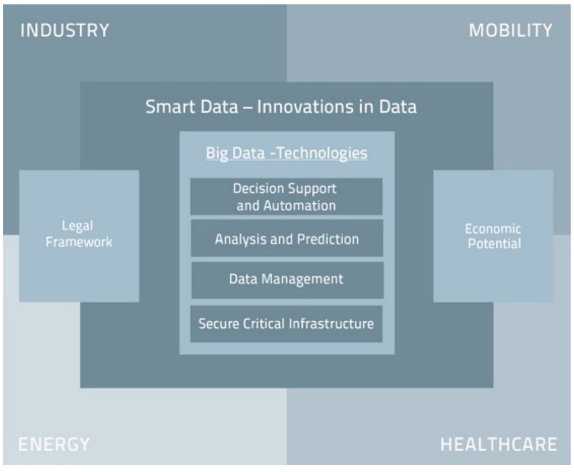

Fig. 2 Overall concept of the "smart data" programme, funded by the German Federal Ministry for Economic Affairs and Energy

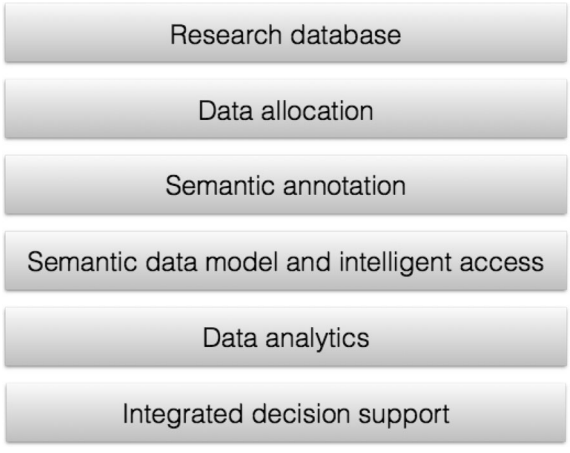

Fig. 3 Work packages

Towards smart data

Expanding the boundaries of health informatics has recently been recognised as one of the top research topic for international conventions on medical care in the twentyfirst century.[3]

Artificial Intelligence (AI) can play a major role and help define the topics in the transition of big data towards smart data, where smart data refers to the meaningful subset of big data. Personalised health data can become the driver of health care innovation and delivery. Specifically, this would be a significant shift from the paradigm where physicians make patient treatment decisions only based on their clinical experience and by evidence-based results derived from general population studies instead of clinical studies. A common definition of evidence-based medicine is the following: "evidence-based medicine requires a bottom-up approach that integrates the best external evidence with individual clinical expertise and patient choice" [45]. The integration of these three components should enable the physician (and patient) to enhance their clinical decisions, so the opportunity for optimal patient quality of life and clinical outcomes can be achieved. KDI focusses on the doctor's decisions (possibly to be combined with other movements such as Participatory Medicine). The KDI project envisions to integrate decision support into the clinical routine in the form of personalised decision support systems for special use cases (Fig. 1). The six work packages, which are explained next, are summarised in Fig. 3.

Research database

A comprehensive view on all patient data in a consistent and integrated way is of great importance. The research database (Fig. 4) contains all data collected from a patient across systems such as diagnoses, laboratory values, radiology images, and pathology results in a single view. Data are collected from as many patients as possible to allow for large-scale analytical methods. The KDI project tries to complement and advance the collection of structured data by also integrating unstructured data (free text), imaging data, and genomic-data. For this, the raw data must be extracted from the source systems (e. g. a PACS system) and normalised using standardised terminologies. As for structured data, classical data warehouse approaches have already been applied successfully to unlock information for secondary purposes such as quality management and research. Unstructured data has to be translated into structured information by text mining. The contents of image data can be described (annotated) by using structured reports such as the BIRADS approach for mammography. Furthermore relevant image features describing specific image contents (such as the "density of a breast") can be automatically extracted; OMICS data can be taken from, e. g. study databases.

Data allocation

It is planned that all available data will be uploaded to the joint research database [35] and made avail-

[3] See, e. g. the AAAI Fall symposium on Expanding the Boundaries of Health Informatics using AI: Making Personalized and Participatory Medicine A Reality, http://www.adventiumlabs.com/HIAI14.

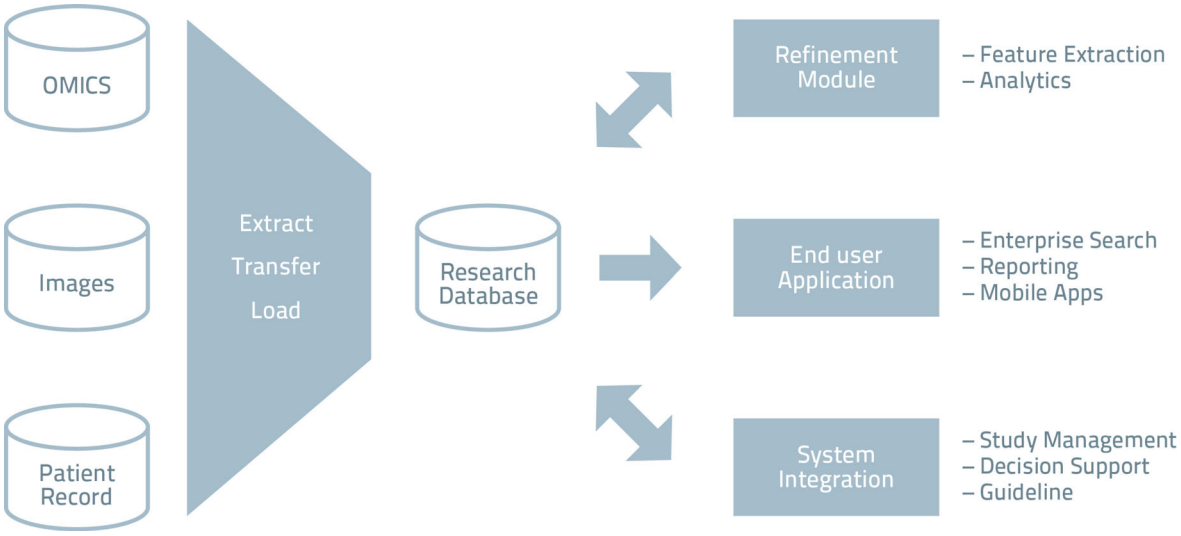

Fig. 4 Research database

able to the project partners for further processing. The infrastructure is based on the tranSMART technology.[4] TranSMART is a data sharing and analytics platform for translational biomedical research. It combines the data-model from the clinical data warehouse i2b2 with the possibility to integrate OMICS data [10, 53]. Major challenges in combining the data are (1) the heterogeneity and completeness of interfaces and semantic data models; (2) the secure and privacy-aware access to data, and (3) the selection and use of appropriate technologies to build a scalable infrastructure. Refinement modules read raw data from the research database, process the data (e. g. machine learning, feature extraction from images or texts) and aggregate the results. Envisioned end user applications may provide physicians and researchers access to the database to search for information or to generate reports. The "smartplatform" is an example of what is possible by a leveraged access to clinical data [32]. The data and results from the refinement modules can be used for multiple clinical and research usage scenarios. Study management systems include patient information; analytical results may support clinical decision making by automatically providing personalised recommendations. All components are installed and operated within the participating clinics Uker and Charité for legal reasons. However, a de-identified (anonymous) shadow database will be implemented to allow access for external developers [13, 19, 40–42].

[4] http://transmartfoundation.org/.

Anonymisation, security, privacy

De-identification of a new set of allocated and relevant patient records is of critical importance. Removing protected health information (PHI) is a critical step in making medical records accessible to more people and automatic procedures of data analytics; here we rely on commercial technology provided by [11].

Many legal and ethical issues arise when processing personal data. Collected patient data may only be used for immediate treatment purposes. Other kinds of processing such as for quality management or research purposes is strictly regulated. This is why removing PHI is a crucial step in making medical records accessible to other kinds of secondary purposes in a de-identified (anonymous) way. Especially free-text data is problematic, as patient-related passages cannot easily be identified. A tool that has been developed in previous projects will thus be further improved to account for heterogeneous unstructured input and thus make narrative reports, discharge letters and other kinds of data available for further analytics.

Semantic annotation

With traditional applications, users may browse or explore visualised patient data such as radiology images, but little to no help is given when it comes to the interpretation of what is being displayed. This

is due to the fact that the semantics of the data are not explicitly stated, hence the semantics therefore remain inaccessible to the system and in turn also to the medical expert using such a system. This can be overcome by incorporating external medical knowledge from ontologies which provide the meaning (i. e. formal semantics) of the data at hand [50].

Texts. Information extraction from text is typically performed by a combination of natural language processing and machine learning approaches. The automatic analysis of medical texts is particularly challenging since sentences are often incomplete, use clinic-specific terms, and contain an abundance of negations. Biomedicine remains one of its most interesting application domains for automatic interpretation of, e. g. a physician letter. This is primarily due to the potentially very broad impact of biomedical findings, but also to the extensiveness of electronic knowledge sources that need to be exploited in an innovative way by integrating natural language processing and machine learning techniques. Current state-of-the-art methods are able to reliably detect key entities in texts; the extraction of statements from texts (much more valuable for describing the patient's status) is still a challenge. So the aggregation of data representation does not necessarily enhance data quality. A feasible solution is that the degree of uncertainty is represented in the annotations and the consecutive processing steps take this uncertainty measure into account.

A recent related example task or automated text mining has been introduced in the 2014 i2b2 NLP challenge: Identifying risk factors for heart disease over time. Medical records for diabetic patients contain information about heart disease risk factors such as high blood pressure and cholesterol levels, obesity, smoking status etc. This challenge aimed to identify the information that is medically relevant to identifying heart disease risk, and track their progression over sets of longitudinal patient records. Within the KDI project, relation extraction from complete sentences for example, is done via a minimally-supervised machine learning system for relation extraction from free text, consisting of two parts, a rule learning and a relation extraction stage feeding each other in a bootstrapping framework, and starting from so-called "semantic seeds" [65]. In addition, we will draw particular attention to the task-oriented extraction of temporal information in the case of "clinical narratives" supporting the structured laboratory data [18, 34, 54, 55, 67].

Images. The automatic analysis of image source data of various modalities (radiology, pathology, EKG, EEG, …) has made great progress in recent years. Since THESEUS MEDICO [48], it is possible to locate and measure major organs as well as healthy and malicious lymph nodes. We expect that our project will greatly benefit from the inclusion of imaging features and semantic image annotations.

The automatic analysis and information extraction from images of various modalities (e. g. radiology, sonography, endoscopy, microscopy) or from biosignals of various types such as ECG, EEG, EMG, has made great progress in recent years. Volumetric image data such as MRI or CT data sets can also be processed [46, 47, 61, 62].

Nevertheless, a complete automated high-level information extraction from medical images is yet an unsolved challenge, which shall partially be addressed within the KDI project using digital and digitised mammographies (radiological projection images from the breast) as an example. Up-to-date many image analysis approaches have been proposed to automatically or interactively detect [27] and delineate mammographic lesions [14] in radiographic breast images. From the delineated image regions a set of descriptive features describing various visible aspects such as contrast, "texturedness", "roughness", "elongation", "size", "compactness" or more complex characteristics such as spectral properties of a lesion are automatically extracted [59, 60]. These image features are then used to train a classifier in order to distinguish "malign" from "benign" lesions [15]. Nevertheless, visible properties such as "texturedness", "coarseness" or "cloudiness" of an image region are quite hard to formulate into computer interpretable form. Instead, it is common practice to use a plethora of different machine vision features, as well as geometric and spectral features from the available images. Using the set of all possible extracted features, automatic selection methods and machine learning approaches are applied to find the best subset of characteristic features for a certain classification and diagnostic problem. Within the KDI project new image features shall be implemented and evaluated in order to address the challenge to identify woman with a higher risk for breast cancer for early detection.

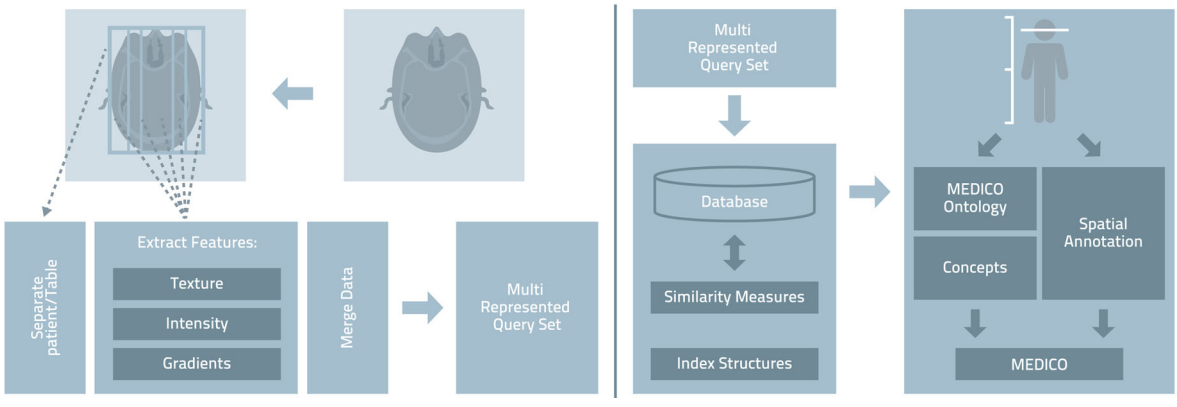

Fig. 5 Scope of the KDI patient data model

Semantic data model and intelligent access

In the healthcare sector, more and more data about the patient's health status as well as about medical background knowledge becomes available. All this patient's data can be stored in different data repositories (e. g. in the various healthcare information systems such as PACS, KIS, HIS, RIS or LIS) in very heterogeneous formats (such as structured lab results or unstructured medical reports or images) and at distributed locations (being the general physician, the hospital, some medical specialist, or the rehab hospital etc.) As of today, only a small percentage of the patient-related data can and is (re)used in advanced clinical applications [69]. This is mainly due to the fact that the available patient data is not semantically integrated and, thus, high efforts for reusing the data sources in advanced clinical application are needed. With the above-mentioned THESEUS MEDICO project several clinical decision support applications were able to demonstrate the high value of semantically integrated patient data. For instance, the reasoning-based application for lymphoma patient classification became possible through the semantic integration of medical image annotation with the related patient data [68]. It has also been demonstrated [38] how semantically integrated patient data can be interpreted in a context-sensitive manner by integrating medical background knowledge to improve clinical decision support. In addition, the semantic annotation of clinical data with concepts or codes from established domain ontologies covering medical and clinical knowledge such as the Foundational Model of Anatomy (FMA), SNOMED CT or the International Classification of Diseases (ICD) needs to be complemented by an ontologically well founded semantic model to structure the references and links to these ontologies. Only if the semantic labels associated with the patient data sources are aligned and used in a consistent manner, a seamless access to heterogeneous patient data becomes possible. This can be realised by means of a semantic patient data model that aligns the various related terminologies and coding systems being used for labelling content.

Today's available models of clinical information such as the HL7 Reference Information Model, however, lack a well defined ontological foundation. In the KDI project, we will try to focus on the development of a semantic data model enabling the seamless access to longitudinal patient data for advanced analytics application (Fig. 5). This task aims to extend the recent research work on establishing a semantic model for clinical information [39].

Medical guidelines

From a technical viewpoint, the clinical state-of-the-art in medical decision making is represented in medical guidelines. Guidelines reflecting common best practice, are based on the results of well understood clinical trials and are at the core of the implementation of evidence based medicine. There have been some efforts to formalise medical guidelines and make them computer-accessible. Computerised clinical guidelines have contributed to the formalisation and automation of clinical data and knowledge. Of particular interest are guidelines to generate medical logic modules [1] because in KDI we interpret smart data as data which can feed personalised decision support systems by reasoning

mechanisms. A first step towards this direction will be to improve understanding how current recommendations are written and to test the adequacy of guideline models [26] towards clinically relevant reasoning and decision support.

Data analytics

Internationally, there are various initiatives under way to improve patient care by collecting and analysing patient and outcome information across providers. An example in the US is the Health Information Technology for Economic and Clinical Health Act (HITECH Act). In this context, large volumes of data are collected and it is hoped that improvements in healthcare can be based on the analysis of this data. A first general result is that fundamental changes in the healthcare system are required and data privacy, data ownership and data security issues must be resolved to make these efforts effective. A more viable solution effective already today would be the analysis of data of individual clinics, as pursued in this project. Thus the goal is not as much to improve processes across clinics but to improve processes in each clinic individually.

From an analytical point of view we analyse data from many individual patients in a complex temporal context with many possible complaints, diagnosis, treatments, procedures and diagnostic results. The basis of our approach is a probabilistic data base model derived from the clinical research data base via machine learning. We use a generalisation of similar approaches for the modelling of knowledge graphs as described in [12, 36, 37]. This probabilistic data base model is based on the learning of latent embeddings and can be the basis for an analysis of the clinical processes but also for the derivation of a number of decision support modules. As discussed in [12, 36, 57], the probabilistic data base model can also support information extraction from unstructured data such as texts and images.

This approach has already been successfully applied to predict hospital readmissions of patients: unscheduled readmission is a well-known problem of healthcare providers around the world and caused additional costs of 17.4 Billion dollars in the US in 2004 and £1.6 Billion in the UK (see USA Hospital Readmission Reduction Program (HRRP); in UK, there is also no reimbursement for within 30 days emergency readmissions).

Readmissions put a high burden not only on the health care system, but also on the patients since complications after discharge generally lead to additional burdens. Our experiments showed that a predictions model that exploits latent embeddings can lead to improved readmission prediction models [3, 8, 56, 66]. A predictive model [28] could also take last lab results, average lab results during the last year, medications prescribed on the previous visit, and medications prescribed during the last year into account.

Integrated decision support

Integrated decision support has two components. First, to answer the question of how to get from guidelines to clinical decision support (for which a unified approach to translating and implementing medical knowledge [33] through semantic annotation is needed), and second, how we can demonstrate that the support is relevant for clinical decisions (as in [9]). Towards this goal, mobile radiology interaction and decision support systems have been discussed recently [52]; in the KDI project, their potential is tried to be extended to data analytics on the large scale by

- Methods for knowledge extraction and personalisation via intelligent predictive analytics;
- Design of personalised care systems to disseminate the discovered knowledge (and enable patients to provide feedback to physicians about their ongoing care);
- Supporting personalised care delivery by modelling patient-focused workflows, supporting their adaptation, and implementing extended facetted search applications.

Currently available clinical data sets represent longitudinal data selected by a physician for the purpose of identifying risk factors. These data can also be used to answer possible other questions on these patients. Some example questions include “Are the medications having the desired effect?”; “Is the patient responding to treatment?”; “Is the patient experiencing an adverse effect from medication X?”

Exploitation

Breast cancer use case

Breast cancer is the most frequently diagnosed solid cancer in women and one of the leading causes

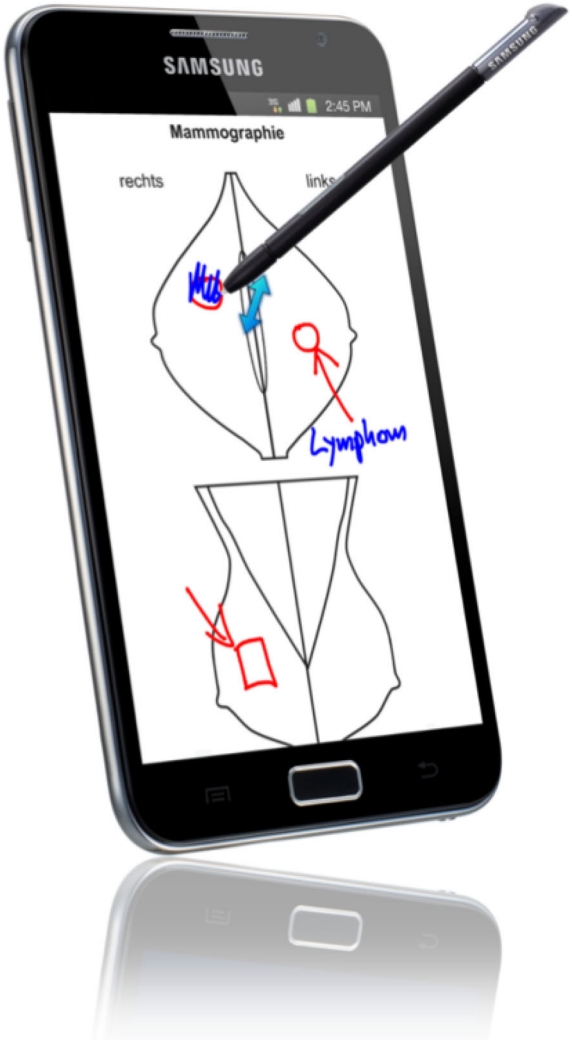

Fig. 6 Integrated mobile medical decision support example

of cancer deaths in the western world [49]. While screening mammography has led to an earlier detection of breast cancer [5], and guideline adherent therapy has improved overall and recurrence-free survival [63], the early detection of breast cancer recurrence remains difficult [44]. Early detection of loco-regional breast cancer recurrence is an important issue because survival is improved if treated adequately [30, 64]. Prediction of the risk to develop breast cancer or cancer recurrence and statements regarding the prognosis remain challenging. A variety of genetic factors and diverse types of information from the clinical examination and different imaging modalities (mammography, sonography, MRI) are of clinical interest. This information is valuable regarding therapy planning and the follow-up duration and intervals. Currently, no reasonable linkage is implemented between the different types of information. We make use of the Bavarian Breast Cancer Cases and Controls database which includes thorough clinical information, more than 1,000,000 genotypes from the germ line DNA and comprehensive image information from all patients and healthy controls. The objective of this use case is to establish a user-friendly application that allows the structured storage, linkage and evaluation of the different data. This provides a unique database for applications in the area of big data technologies and potentially improves patient care and expedites personalised medicine. Figure 6 shows DFKI's intelligent user interface of an integrated decision support system, where automatic labels are combined with human expert annotations based on [2, 17, 22, 24, 43, 58]. It should combine results from (1) textual information extraction, (2) facetted search and (3) medical guidelines. As a result (4), the usability of integrated decision support components should be increased.

Nephrology use case

Kidney diseases are causing a significant financial burden for the German health system. It is estimated that alone the treatment of end-stage renal disease (ESRD) with chronic renal replacement therapies account for more than 2.5 billion euros annually, and the incidence of dialysis-dependent renal insufficiency is rising by 5–8 % each year [31]. Despite progress in diagnosis, prophylaxis and therapy of chronic kidney diseases, renal transplantation remains the therapy of choice for all patients with ESRD. Kidney transplantation leads to a significant improvement of quality of life, to substantial cost savings and most importantly to a significant survival benefit in relationship to other renal replacement therapies. 2272 kidney transplantations were performed in Germany in 2013 but more than 8000 patients are registered on the waiting list for a kidney transplant.[5] The reduction of complications and the increase of long-term graft survival are the main goals after transplantation, against the background of current dramatic organ shortage. It is not only important to reduce or avoid severe or life-threatening complications such as acute rejection,

[5] http://www.dso.de/organspende-undtransplantation/transplantation/nierentransplantation.html.

malignancy and severe opportunistic infections, but it is also of utmost importance to ameliorate the many other serious side effects, which increase cardiovascular risk, decrease renal function, necessitate costly co-medication or hospitalisations and also have an impact on the quality of life after successful transplantation. In many ways, the follow-up after renal transplantation is typical for the supply of chronically ill patients. First, typical complex decisions (allocation of suitable donor organs, selection of different therapeutic regimes etc.) have to be made. Second, the long-term follow-up program offers the possibility to gather long-term histories of patients. This provides a unique data base for applications in the area of big data technologies [4, 6, 7, 23, 25, 29].

Discussion

Automatic procedures, in the sense of automatic decision support systems, bring up the problem of transparency in the knowledge, behaviour and assumption change in the clinical decision process. The clinical decision process is the responsibility of the referring doctor. In addition, the intelligent modulation of granularity of the decision support "feedback" poses a usability problem. KDI includes the testing of the usability of integrated decision support software and evaluation of how easily users learn and use the software to achieve their goals. For example, providing confidence intervals that allow the physician to choose which one of the different answers is most applicable to the situation at hand potentially biases his or her own decision process. In addition, the role of the decision support system must be included in a medical guideline. The starting point is the formalisation of the support of a medical doctor's decision making process as not making decisions for the physicians or the patient for that matter.

In addition, employing increased autonomy of the decision support system poses a problem, especially in clinical application domains, because the doctor's trust in the system may be increased or decreased as a negative side effect. [20] argue that it is difficult for users to trust and rely on complex interaction systems, particularly when the underlying knowledge, behaviour and assumptions of the system are constantly changing and adapting through the use of machine learning.

However, our very strong belief is that having such a powerful tool can provide superior patient care for the individual patient and also strengthen the patient-caretaker relationship.

Conclusion

KDI is the first German medical data intelligence initiative where clinical data is tried to be turned into smart data for clinical decision support. Building a semantic model that can be used to answer treatment-related questions requires human expert knowledge. In this sense, model building requires human expert intervention. In addition to an accurate reasoning model (the semantic net together with reasoning support) another important part is the presentation of the results in specific environmental conditions. Well-integrated decision support may communicate information quickly and can make even complex decisions explorable and understandable at a glance.

Acknowledgements

Thanks go out to the reviewers. This research has been supported by the Smart Data Programme in the KDI project of the Federal Ministry for Economic Affairs and Energy (01MT14001E).

References

1. Agrawal A, Shiffman RN (2001) Using gem-encoded guidelines to generate medical logic modules. In: AMIA 2001, American Medical Informatics Association Annual Symposium, Washington, DC, USA, 3.–7. November 2001, http://knowledge.amia.org/amia-55142-a2001a-1.597057/t-001-1.599654/f-001-1.599655/a-001-1.600134/a-002-1.600131
2. Azzato EM, Tyrer J, Fasching PAEA (2010) Association between a germline OCA2 polymorphism at chromosome 15q13.1 and estrogen receptor-negative breast cancer survival. J Natl Cancer I 102:650–662
3. Barbieri DF, Braga D, Ceri S, Valle ED, Huang Y, Tresp V, Rettinger A, Wermser H (2010) Deductive and inductive stream reasoning for semantic social media analytics. IEEE Intell Syst 25(6):32–41
4. Bissler JJ, Kingswood JC, Radzikowska E, Zonnenberg BA, Frost M, Belousova E, Sauter M, Nonomura N, Brakemeier S, de Vries PJ, Whittemore VH, Chen D, Sahmoud T, Shah G, Lincy J, Lebwohl D, Budde K (2013) Everolimus for angiomyolipoma associated with tuberous sclerosis complex or sporadic lymphangioleiomyomatosis (EXIST-2): a multicentre, randomised, double-blind, placebo-controlled trial. Lancet 381(9869):817–824
5. Bleyer A, Welch HG (2012) Effect of three decades of screening mammography on breast-cancer incidence. N Engl J Med
6. Budde K, Becker T, Arns W, Sommerer C, Reinke P, Eisenberger U, Kramer S, Fischer W, Gschaidmeier H, Pietruck F (2011) Everolimus-based, calcineurin-inhibitor-free regimen in recipients of de-novo kidney transplants: an open-label, randomised, controlled trial. Lancet 377:837–847
7. Budde K, Lehner F, Sommerer C, Arns W, Reinke P, Eisenberger U, Wüthrich RP, Scheidl S, May C, Paulus EMM, Mühlfeld A, Wolters HH, Pressmar K, Stahl R, Witzke O, ZEUS Study Investigators (2012) Conversion from cyclosporine to everolimus at 4.5 months posttransplant: 3-year results from the randomized ZEUS study. Am J Transplant 12(6):1528–1540
8. Bundschus M, Dejori M, Stetter M, Tresp V, Kriegel HP (2008) Extraction of semantic biomedical relations from text using conditional random fields. BMC Bioinformatics 9:207
9. Chaney K, Shiffman RN, Middleton B, White J, Reider J (2013) Findings from a five-year clinical decision support demonstration project and the road ahead. In: AMIA 2013, American Medical Informatics Association Annual Symposium, Wash-

ington, DC, USA, 16.–20. November 2013. http://knowledge.amia.org/amia-55142-a2013e-1.580047/t-04-1.584348/f-004-1.584349/a-056-1.584499/a-065-1.584494
10. Choi IY, Kim TM, Kim MS, Mun SK, Chung YJ (2013) Perspectives on clinical informatics: integrating large-scale clinical, genomic, and health information for clinical care. Genomics Inform 11(4):186–90
11. Daumke P, Enders F, Simon K, Poprat M, Marko K (2012) Semantic Annotation of Clinical Text — the Averbis Annotation Editor. In: Proceedings of the 55th Conference of the German Society of Medical Informatics, Biometry and Epidermiology (GMDS)
12. Dong X, Gabrilovich E, Heitz G, Horn W, Lao N, Murphy K, Strohmann T, Sun S, Zhang W (2014) Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, pp 601–610, ACM, New York, NY, USA, http://doi.acm.org/10.1145/2623330.2623623
13. Dugas M, Lange M, Müller-Tidow C, Kirchhof P, Prokosch H (2010) Routine data from hospital information systems can support patient recruitment for clinical studies. Clin Trials 7(2):183–9
14. Elter M, Held C, Wittenberg T (2010) Contour tracing for segmentation of mammographic masses. Phys Med Biol 55(18):5299–5315
15. Elter M, Schulz-Wendtland R, Wittenberg T (2007) The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process. Med Phys 34:4164–4172
16. Evans WE, Relling MV (2009) Moving towards individualized medicine with pharmacogenomics. Nature 429:464–468
17. Fasching P, Pharoah P, Cox A et al. (2012) The role of genetic breast cancer susceptibility variants as prognostic factors. Hum Mol Genet
18. Gaizauskas RJ, Harkema H, Hepple M, Setzer A (2006) Task-Oriented Extraction of Temporal Information: The Case of Clinical Narratives. In: TIME, IEEE Computer Society, pp 188–195
19. Ganslandt T, Mate S, Helbing K, Sax U, Prokosch HU (2011) Unlocking Data for Clinical Research – The German i2b2 Experience. Appl Clin Inform 2:116–127
20. Glass A, McGuinness DL, Wolverton M (2008) Toward establishing trust in adaptive agents. In: IUI '08: Proceedings of the 13th international conference on Intelligent user interfaces, pp 227–236, ACM, New York, NY, USA, http://doi.acm.org/10.1145/1378773.1378804
21. Groves P, Kayyali B, Knott D, Kuiken SV (2013) The "big data" revolution in healthcare, accelerating value and innovation. In: Centre for US Health System Reform Business Technology Office, Mckinsey & Company
22. Hammon M, Dankerl P, Kramer M, Seifert S, Tsymbal A, Costa MJ, Janka R, Uder M, Cavallaro A (2012) Automated Detection and Volumetric Segmentation of the Spleen in CT Scans. Rofo
23. Hinrichs C, Wendland S, Zimmermann H, Eurich D, Neuhaus R, Schlattmann P, Babel N, Riess H, Gärtner B, Anagnostopoulos I, Reinke P, Trappe RU (2011) IL-6 and IL-10 in post-transplant lymphoproliferative disorders development and maintenance: a longitudinal study of cytokine plasma levels and T-cell subsets in 38 patients undergoing treatment. Transpl Int
24. Hoyer J, Dreweke A, Becker C, Göhring I, Thiel C, Peippo M, Rauch R, Hofbeck M, Trautmann U, Zweier C, Zenker M, Hüffmeier U, Kraus C, Ekici A, Rüschendorf F, Nürnberg P, Reis A, Rauch A (2007) Molecular karyotyping in patients with mental retardation using 100K single-nucleotide polymorphism arrays. J Med Genet 44:629–636
25. Huber L, Naik M, Budde K (2011) Desensitization of HLA-Incompatible Kidney Recipients. New Engl J Med 365(17):1643–1645
26. Hussain T, Michel G, Shiffman RN (2009) The yale guideline recommendation corpus: A representative sample of the knowledge content of guidelines, Int J Med Inform 78(5):354–363
27. Kage A, Elter M, Wittenberg T (2007) An evaluation and comparison of the performance of state of the art approaches for the detection of spiculated masses in mammograms. Conf Proc IEEE Eng Med Biol Soc, pp 3773–3776
28. Krompass D, Esteban C, Tresp V, Sedlmayr M, Ganslandt T (2015) Exploiting latent embeddings of nominal clinical data for predicting hospital readmission. KI – Künstliche Intelligenz, 153–159, http://dx.doi.org/10.1007/s13218-014-0344-x
29. Lasserre J, Arnold S, Vingron M, Reinke P, Hinrichs C (2012) Predicting the outcome of renal transplantation. JAMIA 19(2):255–262
30. Lu W, Jansen L, Post W, Bonnema J, de Velde JV, Bock GD (2009) Impact on survival of early detection of isolated breast recurrences after the primary treatment for breast cancer: a meta-analysis. Breast Cancer Res Treat
31. Lysaght M (2002) Maintenance dialysis population dynamics: Current trends and longterm implications. J Am Soc Nephrol 13:37–40
32. Mandl KD, Mandel JC, Murphy SN, Bernstam EV, Ramoni RL, Kreda DA, McCoy JM, Adida B, Kohane IS (2012) The smart platform: early experience enabling substitutable applications for electronic health records. J Am Med Inform Assoc 19(4):597–603
33. Middleton B, Kawamoto K, Reider J, Rosendale D, Shiffman RN (2012) From guidelines to clinical decision support: a unified approach to translating and implementing knowledge. In: AMIA 2012, American Medical Informatics Association Annual Symposium, Chicago, Illinois, USA, 3–7 November 2012, http://knowledge.amia.org/amia-55142-a2012a-1.636547/t-003-1.640625/f-001-1.640626/a-188-1.640661/a-189-1.640658
34. Mkrtchyan T, Sonntag D (2014) Deep Parsing at the CLEF2014 IE Task (DFKI-Medical). In: CEUR Workshop Proceedings, vol 1180, pp 138–146
35. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, Kohane I (2010) Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc 17(2):124–130
36. Nickel M, Tresp V, Gabrilovich E, Murphy K (2015) Relational machine learning for knowledge graphs. In: Proceedings of the IEEE Conference. IEEE
37. Nickel M, Tresp V, Kriegel HP (2012) Factorizing YAGO: scalable machine learning for linked data. In: Proceedings of the 21st International Conference on World Wide Web Conference, (WWW), pp 271–280. ACM, New York, NY, USA, http://doi.acm.org/10.1145/2187836.2187874
38. Oberkampf H, Zillner S, Bauer B, Hammon M (2012) Interpreting Patient Data using Medical Background Knowledge. In: Proceedings of the International Conference on Biomedical Ontologies (ICBO) 2012, Austria, Graz
39. Oberkampf H, Zillner S, Bauer B, Hammon M (2013) An OGMS-based Model for Clinical Information (MCI). In: Proceedings of International Conference on Biomedical Ontology 2013, pp 97–100
40. Prokosch H, Beck A, Ganslandt T, Hummel M, Kiehntopf M, Sax U, Ückert F, Semler S (2010) IT Infrastructure Components for Biobanking. Appl Clin Inform
41. Prokosch H, Ries M, Beyer A, Schwenk M, Seggewies C, Köpcke F, Mate S, Martin M, Bärthlein B, Beckmann MW, Stürzl M, Croner R, Wullich B, Ganslandt T, Bürkle T (2011) IT Infrastructure Components to Support Clinical Care and Translational Research Projects in a Comprehensive Cancer Center. In: User Centered Networked Health Care – Proceedings of MIE International Congress of the European Federation for Medical Informatics, Oslo, Norway
42. Prokosch HU, Ganslandt T (2009) Perspectives for medical informatics. Reusing the electronic medical record for clinical research. Method Inform Med 48:38–44
43. Rauch A, Thiel C, Schindler D, Wick U, Crow Y, Ekici A, van Essen A, Goecke T, Al-Gazali L, Chrzanowska H, Zweier C, Brunner H, Becker K, Curry C, Dallapiccola B, Devriendt K, Dörfler A, Kinning E, Megarbane A et al (2008) Mutations in the pericentrin (PCNT) gene cause primordial dwarfism. Science 319:816–819
44. Rojas M, Telaro E, Russo A, Moschetti I, Coe L, Fossati R, Palli D, del Roselli T, Liberati A (2005) Follow-up strategies for women treated for early breast cancer. Cochrane Database Syst Rev
45. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS (1996) Evidence based medicine: what it is and what it isn't. BMJ 312(7023):71–72
46. Seifert S, Barbu A, Zhou SK, Liu D, Feulner J, Huber M, Sühling M, Cavallaro A, Comaniciu D (2010) Hierarchical parsing and semantic navigation of full body CT data. In: Proceedings of the SPIE Medical Imaging. http://www5.informatik.uni-erlangen.de/Forschung/Publikationen/2009/Seifert09-HPA.pdf
47. Seifert S, Thoma M, Stegmaier F, Hammon M, Kramer M, Huber M, Kriegel HP, Cavallaro A, Comaniciu D (2011) Combined semantic and similarity search in medical image databases. In: SPIE Medical Imaging
48. Seifert S, Zillner S, Huber M, Sintek M, Sonntag D, Cavallaro A (2011) Theseus Usecase MEDICO (in German). In: Acatech diskutiert „Internet der Dienste" (Internet of Services). Springer
49. Siegel R, Naishadham D, Jemal A (2012) Cancer statistics. CA Cancer J Clin
50. Sonntag D, Möller M (2009) Unifying semantic annotation and querying in biomedical image repositories. In: Proceedings of International Conference on Knowledge Management and Information Sharing (KMIS)
51. Sonntag, D., Wennerberg, P., Buitelaar, P., Zillner, S.: Cases on Semantic Interoperability for Information Systems Integration: Practices and Applications, chap. Pillars of Ontology Treatment in the Medical Domain, pp 162–186. Information Science Reference (2010)
52. Sonntag D, Zillner S, Ernst P, Schulz C, Sintek M, Dankerl P (2014) Mobile radiology interaction and decision support systems of the future. In: Wahlster W, Grallert HJ, Wess S, Friedrich H, Widenka T (eds) Towards the Internet of Services: The THESEUS Research Program, Cognitive Technologies. Springer International Publishing, pp 371–382
53. Sreenivasaiah PK, Kim do H (2010) Current trends and new challenges of databases and web applications for systems driven biological research. Front Physiol 1:147

54. Styler WF, Bethard S, Finan S, Palmer M, Pradhan S, de Groen PC, Erickson B, Miller T, Lin C, Savova G, Pustejovsky J (2014) Temporal annotation in the clinical domain. T Assoc Comput Linguist 2:143–154
55. Sun W, Rumshisky A, Uzuner O (2013) Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. J Am Med Inform Assoc 20(5):806–813
56. Tresp V, Huang Y, Nickel M (2014) Querying the Web with Statistical Machine Learning. In: Wahlster W, Grallert HJ, Wess S, Friedrich H, Widenka T (eds) Towards the Internet of Services: The THESEUS Research Program, Cognitive Technologies. Springer International Publishing
57. Tresp V, Zillner S, Costa MJ, Huang Y, Cavallaro A, Fasching PA, Reis A, Sedlmayr M, Ganslandt T, Budde K, Hinrichs C, Schmidt D, Daumke P, Sonntag D, Wittenberg T, Oppelt PG, Krompass D (2013) Towards a new science of a clinical data intelligence. In: Proceedings of the NIPS Workshop on Machine Learning for Clinical Data Analysis and Healthcare
58. Untch M, von Minckwitz G, Konecny GE, Conrad U, Fett W et al., CK (2011) PREPARE trial: a randomized phase III trial comparing preoperative, dose-dense, dose-intensified chemotherapy with epirubicin, paclitaxel, and CMF versus a standard-dosed epirubicin–cyclophosphamide followed by paclitaxel with or without darbepoetin alfa in primary breast cance-outcome on prognosis. Ann Oncol:1999–2006
59. Wagner F, Wittenberg T (2011) New features for the classification of mammographic masses. Int J Comput Appl 35(4):29–35
60. Wagner F, Wittenberg T, Elter M (2010) Classification of mammographic masses: influence of regions used for feature extraction on the classification performance. Proc. SPIE, Medical Imaging
61. Wels M, Kelm BM, Hammon M, Jerebko AK, Sühling M, Comaniciu D (2012) Data-driven breast decompression and lesion mapping from digital breast tomosynthesis. MICCAI (1):438–446
62. Wels M, Kelm BM, Tsymbal A, Hammon M, Soza G, Sühling M, Cavallaro A, Comaniciu D (2012) Multi-stage osteolytic spinal bone lesion detection from CT data with internal sensitivity control. In: SPIE Medical Imaging
63. Woeckel A, Kurzeder C, Geyer V, Novasphenny I, Wolters R, Wischnewsky M, Kreienberg R, Varga D (2010) Effects of guideline adherence in primary breast cancer – a 5-year multi-center cohort study of 3976 patients. Breast
64. Woeckel A, Kreienberg R (2008) First Revision of the German S3 Guideline "Diagnosis, Therapy, and Follow-Up of Breast Cancer". Breast Care
65. Xu F, Uszkoreit H, Li H, Adolphs P, Cheng X (2014) Domain-adaptive relation extraction for the semantic web. In: Wahlster W, Grallert HJ, Wess S, Friedrich H, Widenka T (eds) Towards the Internet of Services: The THESEUS Research Program, Cognitive Technologies. Springer International Publishing, pp 289–297
66. Yu K, Chu W, Yu S, Tresp V, Xu Z (2006) Stochastic Relational Models for Discriminative Link Prediction. In: Advances in Neural Information Processing Systems (NIPS 2006). MIT Press
67. Zhou L, Friedman C, Parsons S, Hripcsak G (2005) System architecture for temporal information extraction, representation and reasoning in clinical narrative reports. AMIA Annu Symp Proc, pp 869–873
68. Zillner S (2010) Reasoning-Based Patient Classification for Enhanced Medical Image Annotations. In: Proceedings of the Extended Semantic Web Conference, (ESWC 2010), Heraklion, Greece, June
69. Zillner S, Neururer S (2015) Technology roadmap for big data healthcare applications. KI – Kuenstliche Intelligenz 29(2):131–141