

# SMEYEDAT (SMART-EYE-DATA): ZUSAMMENFÜHRUNG UND NUTZBARMACHUNG OPHTHALMOLOGISCHER DATEN.

Nasseh D<sup>1</sup>, Müller M<sup>2</sup>, Ahlborn B<sup>1</sup>, Kortüm K<sup>2</sup>, Kampik A<sup>2</sup>, Mansmann U<sup>1</sup>, Kreutzer T<sup>2</sup>  
 (1) IBE, Ludwig-Maximilians-Universität München, Germany  
 (2) Augenklinik der LMU, Klinikum der Universität München



## Einleitung und Fragestellung

Trotz zunehmender Digitalisierung stellt sich die wissenschaftliche Auswertung von Daten in der klinischen Medizin als nicht triviale Aufgabe dar. Bereits auf unterster Ebene, sprich, in den einzelnen Fachabteilungen eines Klinikums werden Daten eines Patienten in verschiedensten Systemen archiviert. Interne Datenbanken zu diagnostischen Spezialgeräten, Verwaltungsdaten, die in Klinikums Informationssystemen gespeichert werden, als auch isolierte Arztbriefe sind hier nur einige der in der Praxis genutzten Entitäten. Oftmals verbirgt sich wissenschaftlicher Mehrwert jedoch gerade in der kombinierten Analyse verschiedener Datenquellen [1].

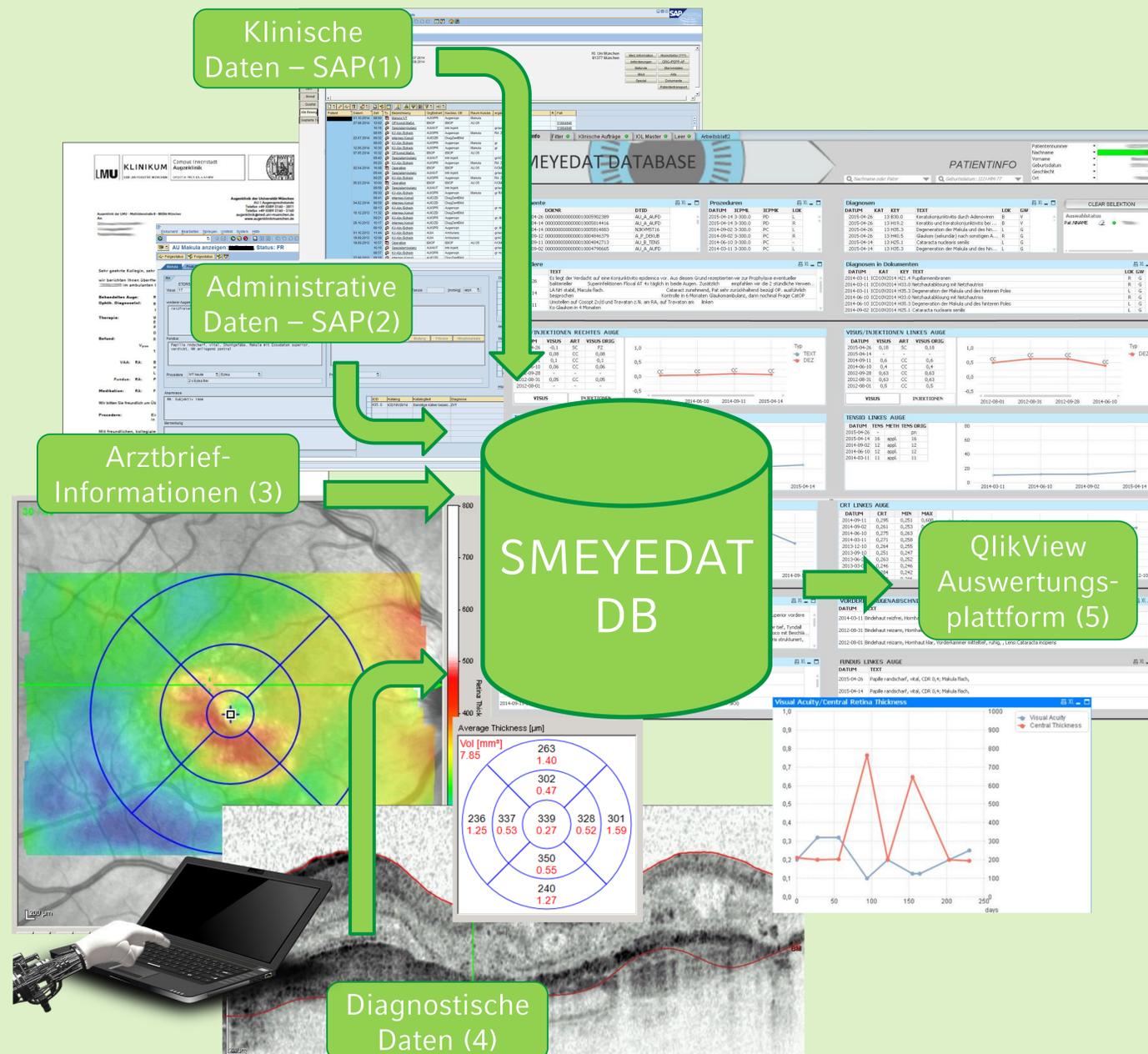
Durch die Hohe Komplexität bzw. Dimensionalität ist hier oftmals von Big-Data die Rede [2]. Auch im Bereich der Augenheilkunde existiert diese Problematik. Zudem hat hier die Bildgebung in den letzten Jahren deutlich an Bedeutung gewonnen und es bestehen nun zahlreiche Möglichkeiten Befunde zu objektivieren und zu messen. In der Augenklinik der Universität in München sind in den letzten Jahren deswegen Anstrengungen unternommen worden, zum einen die klinischen Untersuchungsdaten digital zu erfassen und zum anderen die Messdaten geordnet zu speichern. Das SMEYEDAT (Smart-Eye-Data) Projekt, eine Arbeitsgruppe bestehend aus Ärzten, Informatikern als auch Statistikern, hat sich zur Aufgabe gemacht, diese Datenquellen zusammenzuführen, zu strukturieren und damit für die Forschung nutzbar zu machen.

## Methoden (1)

**Datenzusammenführung:** Zentraler Punkt der Arbeitsgruppe ist die Schaffung einer strukturierten Forschungsdatenbank, initial bestehend aus nachfolgenden Quelldaten:

- Klinische sowie administrative Basisdaten:** Die Datenbasis stammt aus dem KIS des Klinikums der Ludwig-Maximilians-Universität München, welches zur Erfassung und Verwaltung der Daten ein SAP R3 System mit integrierter i.s.h.med Technologie (Siemens AG) verwendet. Die relevanten Informationen (wie patientenbezogene medizinische oder verwaltungstechnische Daten) werden in Form von Textdateien exportiert und unter Berücksichtigung der SAP-internen Identifikationsmerkmale in die SMEYEDAT DB integriert.
- Semi-strukturierte Arztbrief Informationen:** Arztbriefe (im rtf oder doc-Format), deren relevanter Inhalt durch Verwendung eines C#-Parserskriptes in die SMEYEDAT DB überführt wurde. Dabei wurden Abweichungen von der vorgegebenen Struktur der Dokumente bis zu einem gewissen Grad toleriert.
- Diagnostische Daten (Optische Kohärenztomografie):** Der initiale Datenbestand wurde durch Daten aus dem OCT ergänzt, hierbei sind von besonderer Relevanz, die Netzhautdickewerte der Makula, also dem Bereich des schärfsten Sehens. Die Daten werden hierbei in Echtzeit beim Aufruf eines OCTs generiert und liegen nicht in abgespeicherter Form vor. Aktuell ist es über die gegebene Software des Herstellers nicht möglich die [...]

## Zusammenfassende Darstellung



**Abbildung:** Die SMEYEDAT-DB ist ein Zusammenschluss aus Daten verschiedener Datenquellen die für die wissenschaftliche Auswertung im Bereich der Ophthalmologie nutzbar gemacht werden. Die zugrunde liegenden Daten stammen bisher aus dem klinischen Informationssystem des Klinikum der Universität München und umfassen hierbei sowohl klinische Daten (1) als auch administrative Daten (2). Zusätzlich wird die SMEYEDAT-DB durch semi-strukturierte Informationen aus Arztbriefen ergänzt (3). Final wurde der Datensatz durch schwer zugängliche Daten aus diagnostischen Geräten (OCT), die mittels eines robotischen Skriptes fallweise einzeln exportiert werden, dem Datensatz hinzugefügt (4). Die SMEYEDAT-DB ist hierbei nicht nur eine lose Datensammlung sondern eine strukturierte Datenbank, auf der eine Instanz von Qlik-View (5) aufgesetzt wurde. Die Auswertungs-plattform erlaubt es schnell und über wenige Anweisungen die gegebenen Daten für Forschungszwecke oder unterstützend zur Routine auszuwerten.

## Methoden (2)

- [...] gewünschten Daten im Batch gesammelt zu exportieren. Für den Export wurde ein in Java implementiertes Roboterskript erzeugt, das extern über Mauskommandos und Tastatureingaben die Hersteller Software betätigt, die einzelnen Daten jeweils aufruft, in Echtzeit generiert und schließlich als XML-Datei exportiert. Über Bilderkennung passt sich das Roboterskript hierbei jeweils an die zugrundeliegenden Daten an und kann entsprechend den Export durchführen. Die Vollständigkeit der Daten wurde nachträglich validiert.

**Synchronisation:** Für die gegebenen Datenquellen wurden Methoden entwickelt, die den Datenbestand der SMEYEDAT DB aus den Quellsystemen regelmäßig, vollautomatisch aktualisiert.

**Datenauswertung:** Die Daten werden mit Hilfe der grafischen Analyse-Software QlikView ausgewertet.

## Ergebnisse

(Stand Februar 2015)

- Aus dem Fachgebiet übergreifenden KIS des KUM wurden Daten von über 250.000 Patienten integriert
- Informationen aus mehr als 7.000 verschiedenen Arztbriefen
- Diagnostische-Daten zu 7.852 Patienten mit mehr als 33.651 Untersuchungen (OCT-scans, Angiographien etc.)
- Strukturiertes Datenmodell der SMEYEDAT-DB
- QlikView Auswertungsplattform

## Diskussion

Im Gegensatz zu Ansätzen die eine Datenzusammenführung und Auswertungen innerhalb einer Datawarehouse-Plattform durchführen (z.B. I2B2 [3]) wurde ein strukturiertes Datenmodell auf MS-SQL Basis erarbeitet um Echtzeitauswertungen über die grafische Analyse-Software QlikView, die auf diesen Datenbestand zugreift, zu ermöglichen und so auch den praktizierenden Ärzten ein Tool zur Verfügung zu stellen, Analysen wie beispielsweise Krankheitsverläufe auf großen Datenbeständen selbstständig zu modellieren. Technische Probleme beim Import, besonders beim Auftreten von Spezialfällen (z.B. Patienten mit extrem vielen Diagnosedaten), ließen sich fast vollständig lösen. Inhaltliche Probleme wie beispielsweise eine Verbesserung der Trennungsschärfe von Diagnosen müssen Anforderungsbezogen nachbearbeitet werden. Aus den vorliegenden Informationen können jedoch bereits jetzt umfassende medizinische Verlaufsmuster wie beispielsweise die Entwicklung der Visuswerte (Sehqualität) oder des Augeninnendruckes eines Patienten gewonnen werden.

## Referenzen

- [1] Lluch-Ariet M, Brugués de la Torre A, Vallverdú F, Pegueroles-Vallés J. Knowledge sharing in the health scenario. *J Transl Med.* 2014 Nov 28;
- [2] Panahiazar M, Taslimitehrani V, Jadhav A, Pathak J. Empowering Personalized Medicine with Big Data and Semantic Web Technology: Promises, Challenges, and Use Cases. *Proc IEEE Int Conf Big Data.* 2014 Oct;2014:790-795.
- [3] Segagni D, Tibollo V, Dagliati A, Malovini A, Zambelli A, Napolitano C, Priori SG, Bellazzi R. Clinical and research data integration: the i2b2-FSM experience. *AMIA Jt Summits Transl Sci Proc.* 2013 Mar 18;2013:239-40. eCollection 2013.