

Perspektiven der pharmazeutischen Industrie auf Big Data Analysen

Friedhelm Leverkus
Director HTA&OR



**DIESE PRÄSENTATION STELLT NUR DIE
SICHTWEISEN DER VORTRAGENDEN UND NICHT
DIE POSTION DES UNTERNEHMES PFIZER DAR.**

- Beispiele: Cancer Link und IBM Watson
- Versorgungsforschung und Big Data
- Datenschutz



Patientendaten

Der Schlüssel für Forschung und Versorgung



Produktpipeline von Pfizer, Stand: 02. August 2016

Pfizer Produktpipeline

Forschung und Entwicklung

Drei große Trends wirken zusammen und revolutionieren die zukünftigen Nutzungsmöglichkeiten von aggregierten Daten für die Behandlung von Krebspatienten

1. “genomics” und weitere so genannte “omics” (metabolomic, proteomic und microbiomic) verändern das Verständnis der Pathogenese von Krebs
2. Zunehmende Verbreitung von elektronischen Patientenakten (electronic health records - EHRs) und wachsende Fülle von digitalen Informationsquellen über Patienten, inkl. Patient-reported Outcomes, Daten von tragbaren Sensoren (Wearables)
3. Bisherige Grenzen der Analysemöglichkeiten erweitern sich und ermöglichen den Umgang mit zunehmenden Datenmengen. Analytische Ansätze wie “machine learning” und neue statistische Methoden versprechen die Aufdeckung von bisher unerkannten Verbindungen und steigern den Wissenszuwachs

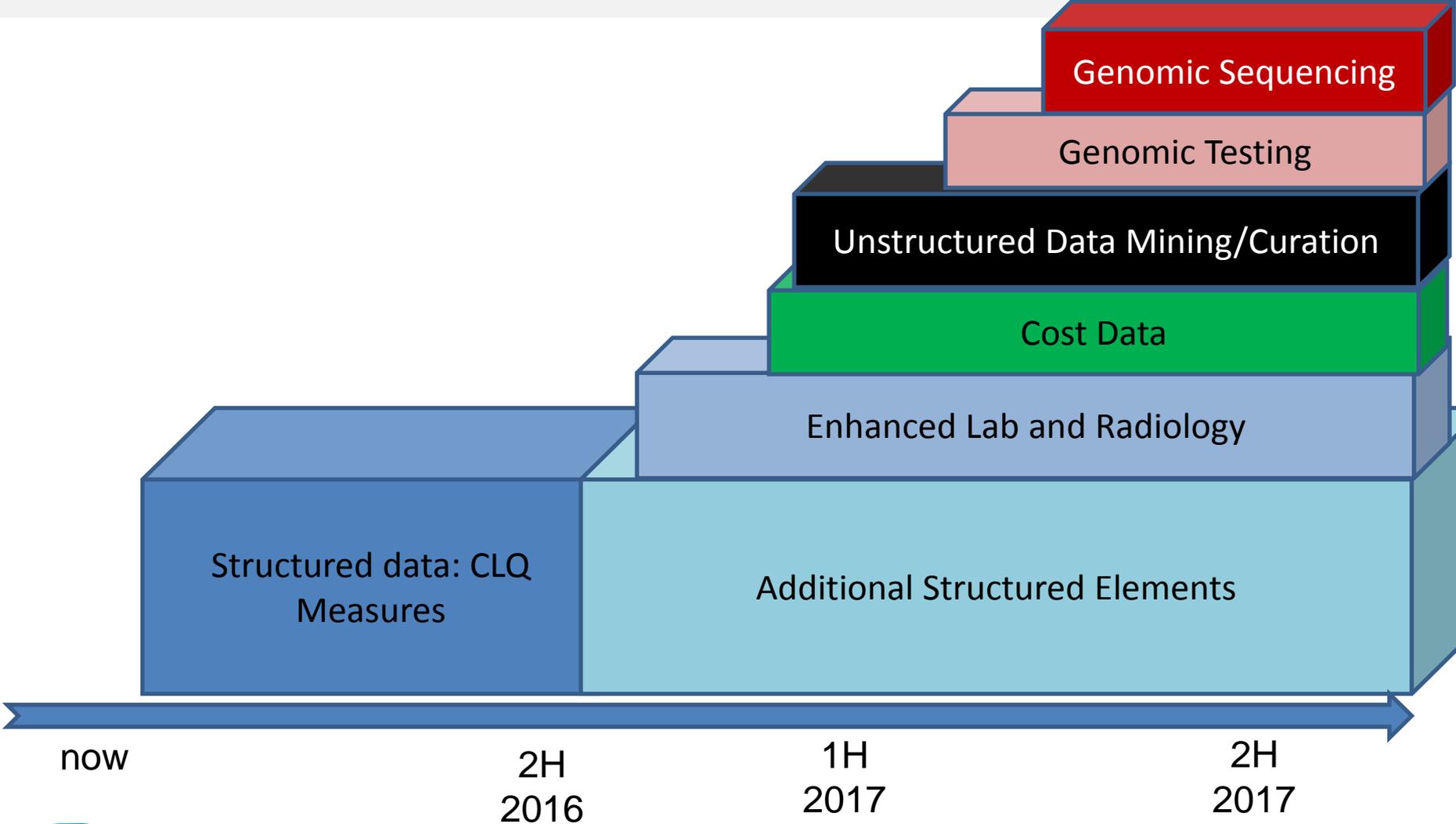


Shaping the future of cancer care

ASCO's CancerLinQ aims to rapidly improve the overall quality of cancer care, and is the only major cancer data initiative being developed and led by physicians. When complete, CancerLinQ will unlock real-world patient care data from millions of electronic health records, and securely process and analyze the data to provide immediate quality feedback and clinical decision support to providers. Doctors will receive personalized insights on a scope that was previously unattainable, and patients will benefit by having access to high quality care based on up-to-date insights and findings.

- CancerLinQ wurde 2011 von der ASCO konzipiert und gegründet, um das Lernen der Ärzte und die Qualität der Versorgung zu verbessern
 - Finanzierung durch ASCO und philanthropische Spenden
- November 2014 – Neustrukturierung als eigenständiges Unternehmen mit CEO und Personal, um das Projekt voranzutreiben
- Januar 2015 – SAP kommt als strategischer Partner hinzu
 - Stellt als technische Grundlage SAP Hana Plattform zur Verfügung
 - Zugang zu Programmen, Innovationslabor
 - Globale Synergien
- Neuer CEO eingesetzt im August 2015
- Herbst 2015: Beginn der Verbindungen mit der onkologischen Community

CancerLinQ data map



CancerLinQ Clinical User Portal

ASCO CANCERLINQ CONNIEB

My Patients

- Quality Performance Indicators**
Explore your clinical quality measures
25 Measures
- CancerLinQ Insights - My Patients**
Cohort creation and analysis
1,502 Patients
- My Priority Patients**
Take action with your patients
- My Patients**
Search/Sort/Filter individuals in your patient list
- My Case Mix**
Who are my patients?
- Stage IV Colorectal Patients**
Chemotherapy toxicity

Clinical Analytics

- CancerLinQ Insights - All Patients**
Cohort creation & hypothesis generation
1M Patients
- Patient Age Distribution by Canc...**
- Stage IV Bladder Cancers**
Targeting clinical trials
- High Emetogenic Risk Chemotherapy...**
Missed opportunities

Practice Analytics

- All Patient Case Mix**

CancerLinQ will place practical insights in the hands of:



Doctors—Personalized guidance on treatment decisions by matching each patient's data against clinical guidelines and recognized quality standards with the outcomes of patients like them

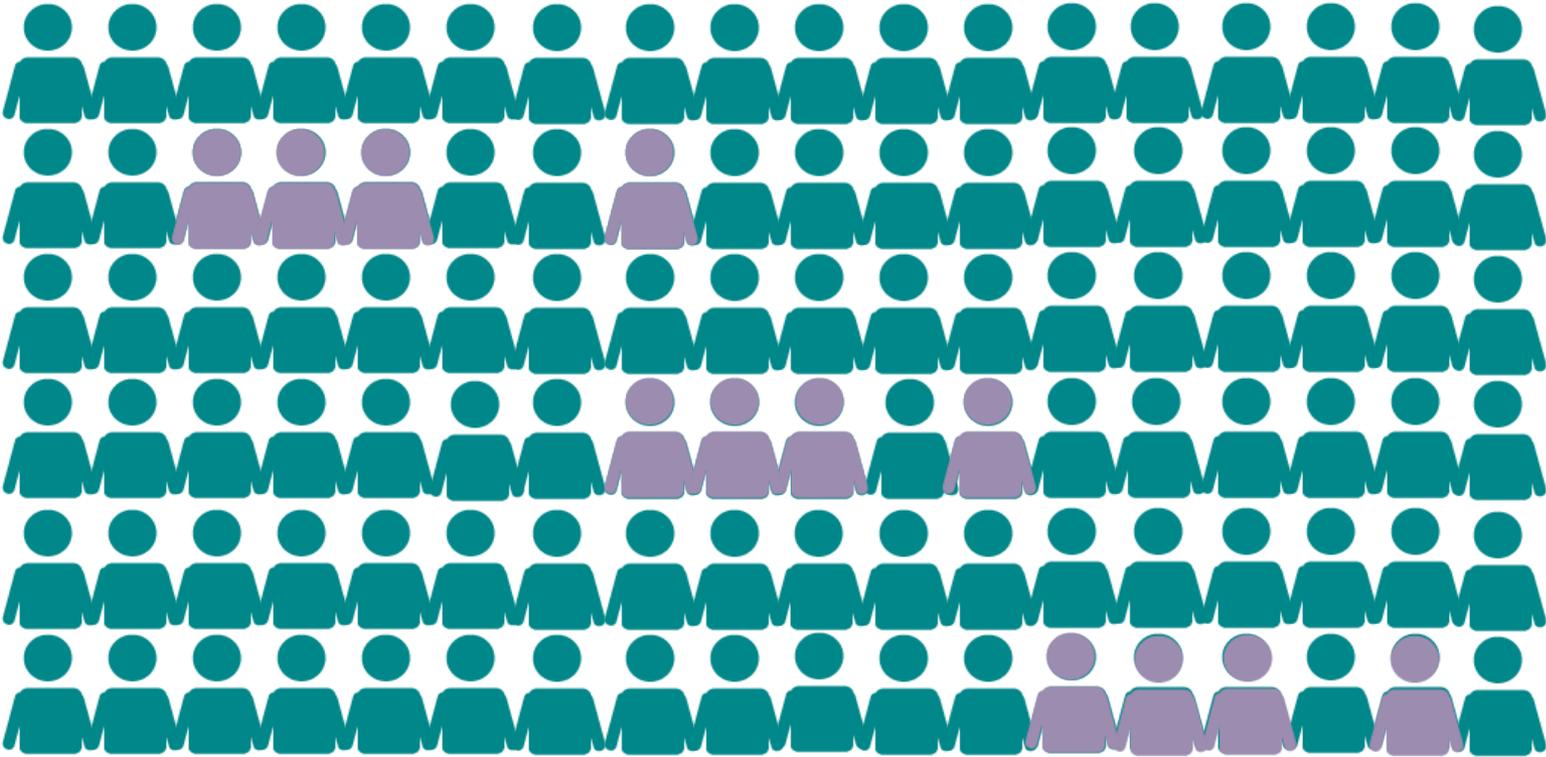


Patients—Reassurance that they're getting high-quality, evidence-based care along with personalized treatment information and better coordination among members of their healthcare team



Researchers—New insights for discovery through access to a massive body of de-identified patient care data to analyze patterns and generate new hypotheses

We could match similar patients in real-time



NCT SAP HANA® Technical Implementation

Key features deployed

- Data services
- Semantic text analytics
- Predictive analytics

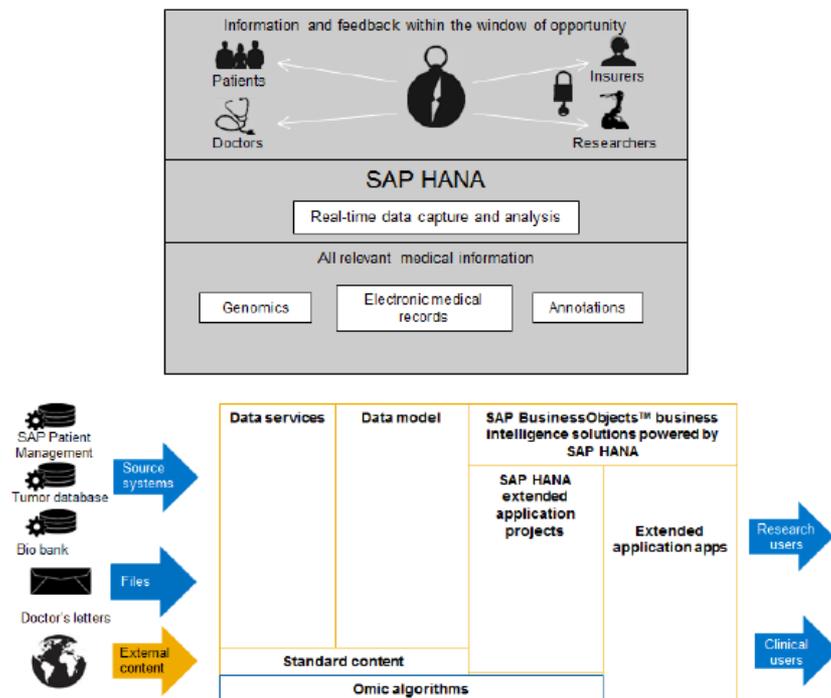
Technical key performance indicators

- Analyze all data sources effectively
- Deploy modern methods of text analysis and knowledge sharing
- Identify biomarkers to support decisions for all aspects of cancer treatment
 - Prognosis: Likelihood to develop cancer
 - Diagnosis: Type of cancer
 - Prediction: Optimal expected response to treatment
 - Medication: Optimal dose for patient (personalized)
 - Recurrence: Likelihood of the cancer returning
 - Clinical trial: Patient cohorts

Partner involvement

- SAP® Consulting organization
- Co-innovation project with SAP
- Proof-of-concept test completed in six weeks, followed by the successful launch of the Medical Research Insights project

NCT SAP HANA® platform architecture



NCT SAP HANA® Technical Implementation

Key features deployed

- Data services
- Semantic text analytics
- Predictive analytics

Technical key performance indicators

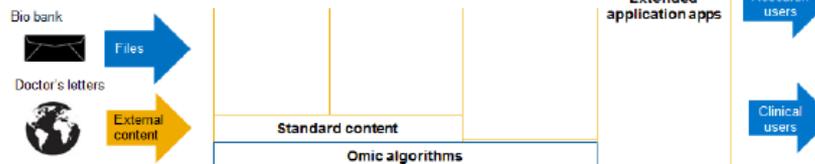
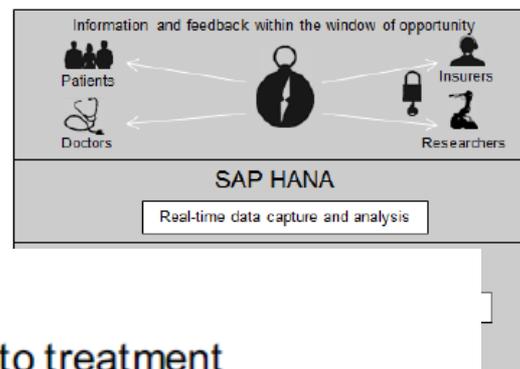
- Analyze all data sources effectively
- Deploy modern methods of text analysis and knowledge sharing

- Diagnosis: Type of cancer
- Prediction: Optimal expected response to treatment
- Medication: Optimal dose for patient (personalized)
- Recurrence: Likelihood of the cancer returning

Partner involvement

- SAP® Consulting organization
- Co-innovation project with SAP
- Proof-of-concept test completed in six weeks, followed by the successful launch of the Medical Research Insights project

NCT SAP HANA® platform architecture



CancerLinQ

CancerLinQ will:

- 1** Measure quality and performance

- 2** Provide data that will be used to continually improve the performance of providers and patient outcomes

- 3** Provide physicians with clinical decision support tools

- 4** Provide insights for exploration and hypothesis generation



WHY IBM WATSON HEALTH?

1



IBM Watson Health is pioneering a new partnership between humanity and technology. Currently, **80% of health data is invisible** because it's unstructured or inaccessible.



Watson technology sees this data, reading **40 million documents in 15 seconds**, while managing one of the largest clinical data sets in the world.

IBM Watson Health uses cognitive health technology to help transform the way organizations deliver health and wellness, harnessing previously untapped medical data, and discovering more than we ever thought possible.



IBM Watson und Pfizer wollen die immunonkologische Forschung beschleunigen

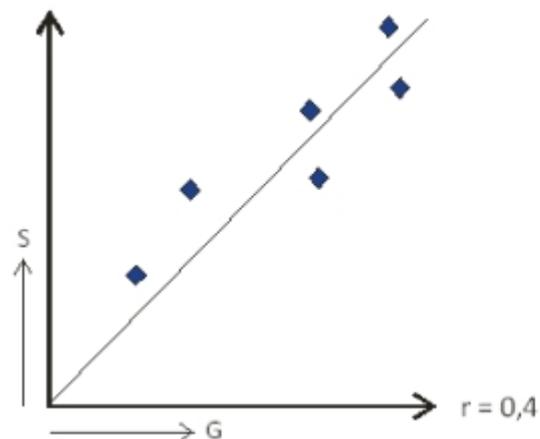
- **IBM Watson for Drug Discovery,**
 - ein cloud-basiertes kognitives Programm
 - maschinelles Lernen
 - Verarbeitung natürlicher Sprache sowie andere kognitive Eigenschaften werden genutzt
 - Watson for Drug Discovery kann Wissenschaftler bei der Analyse von Daten aus unterschiedlichen Quellen unterstützen und durch eine dynamische Visualisierung versteckte Muster und Zusammenhänge erkennbar machen.
 - Beschleunigung des Forschungsprogramms

Implikationen von Big Data im Gesundheitswesen sind andere als bei Amazon



Correlation is not causation

Geburtenrate / Storchendichte



S = Storchendichte (V_{a1})

G = Geburtenrate (V_{a2})



Journal of Clinical Epidemiology 59 (2006) 964–969

Journal of
Clinical
Epidemiology

Testing multiple statistical hypotheses resulted in spurious associations:
a study of astrological signs and health

Peter C. Austin^{a,b,c,*}, Muhammad M. Mamdani^{a,d}, David N. Juurlink^{a,e,f}, Janet E. Hux^{a,c,e,f}

Results: We tested these 24 associations in the independent validation cohort. Residents born under Leo had a higher probability of gastrointestinal hemorrhage ($P = 0.0447$), while Sagittarians had a higher probability of humerus fracture ($P = 0.0123$) compared to all other signs combined. After adjusting the significance level to account for multiple comparisons, none of the identified associations remained significant in either the derivation or validation cohort.

Conclusions: Our analyses illustrate how the testing of multiple, non prespecified hypotheses increases the likelihood of detecting implausible associations. Our findings have important implications for the analysis and interpretation of clinical studies. © 2006 Elsevier Inc. All rights reserved.

Big Data Analytics

Klassische Statistik

- A-priori-Annahmen über Verteilungen und Parameter
- Ein Algorithmus schätzt die vorgegebenen Parameter

Maschinelles Lernen ist auch ein statistische Modell

- Man muss sich viel weniger um die Modellvoraussetzungen kümmern
- Minimale Anforderungen an die Modellstruktur
- „aus den Daten“ einer Maschine (Modell) zu lernen: die Form der Beziehungen wird während des Lernprozesses bestimmt.

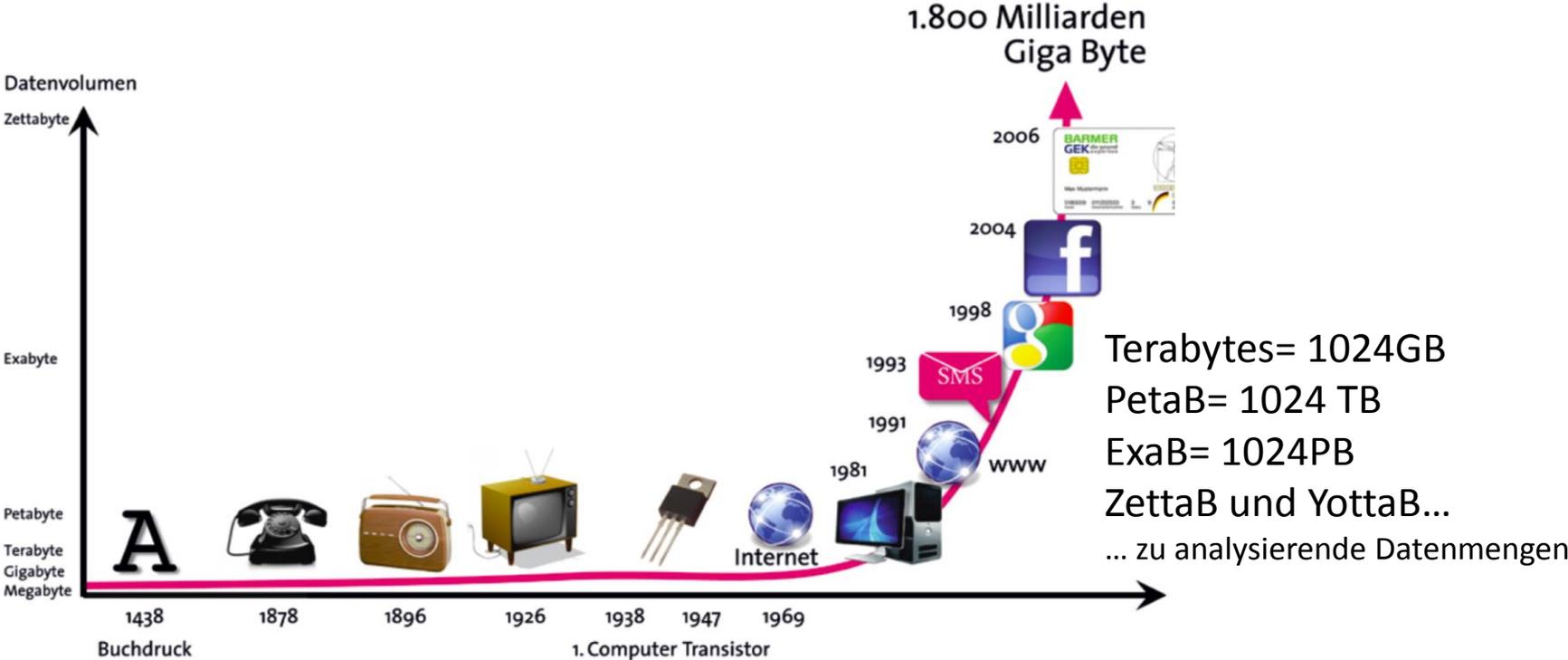
→ **Big Data Analytics Statistik zentral Disziplin**

→ **Studium: Data Scientist mit IT und Statistik Elemente**

→ **Hypothesen Generierung vs Hypothesen Testen**

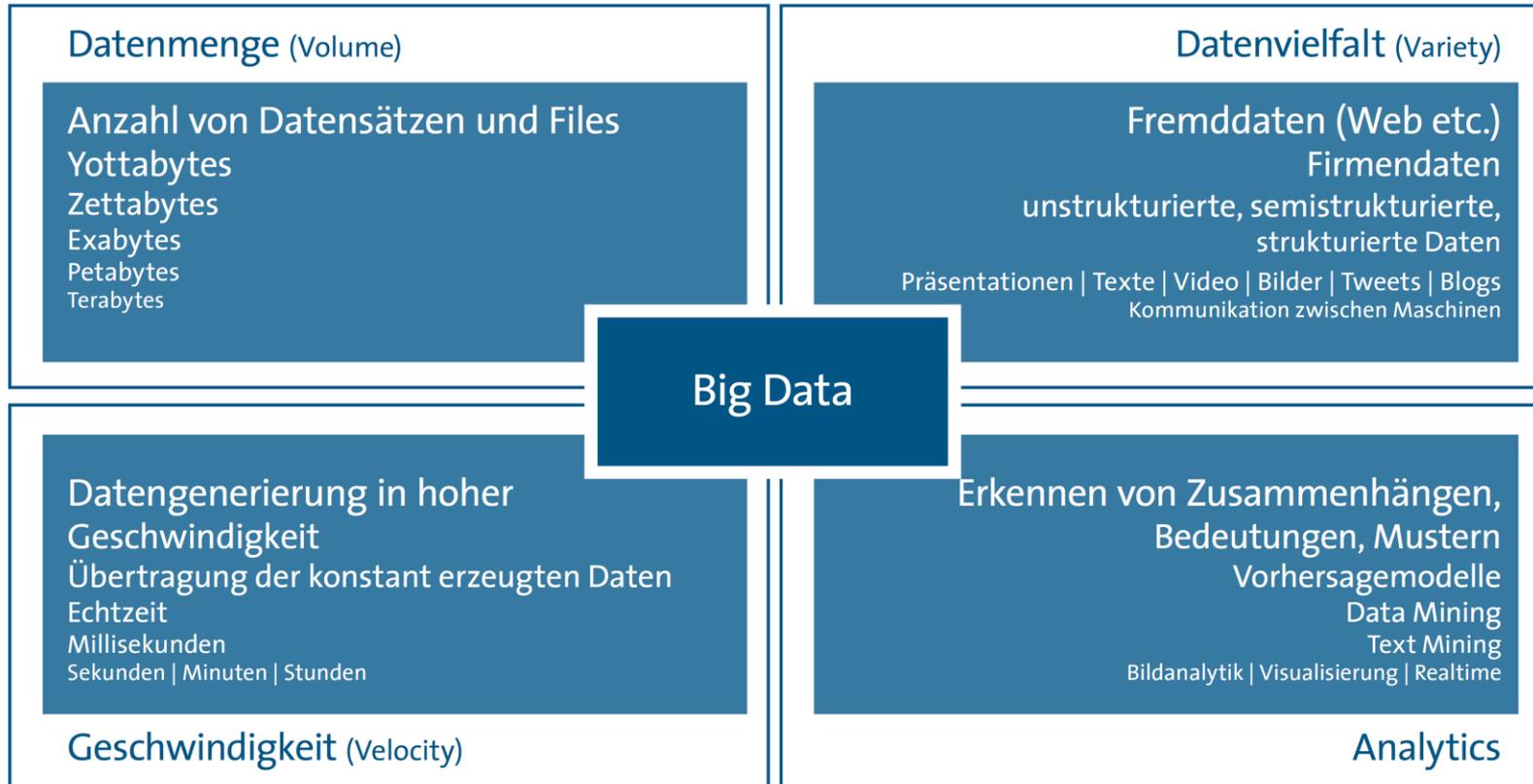
Big Data

Der Begriff “Big Data” entwickelte sich, als Antwort auf die Datenexplosion der letzten Dekade .



¹Jonathan Stuart Ward, Adam Barker Undefined By Data: A Survey of Big Data Definitions <http://arxiv.org/abs/1309.5821>

Big Data Definition



Fragestellung der Versorgungsforschung



- Schätzung von Prävalenz und Inzidenz
- Krankheitskosten
- Über-/ Unterversorgung
- Leitliniengerechte Behandlung
- Qualitätssicherung
- Wirksamkeit und Verträglichkeit unter Alltagsbedingungen
- Data Mining- Identifikation von Patientengruppen mit speziellen Eigenschaften

Was sind RWD und Big Data?

Welche Datentypen werden unter RWD subsummiert?

Ein breites Spektrum an Datenquellen fällt unter den Begriff RWD:

- **Versicherungsdaten**
- **Klinische Daten – strukturiert und unstrukturiert**
 - Krankenakten, Notizen des Arztes, Befunde, Patientenregister, Genomische Datenbanken, Labore, Bilder, Heim- und Patientenbett Monitore und Sensoren
- **Data Linkage**
- **Nationale Kohorte**
- **Reported Data**
 - Social Media, Apps, Blogs, Chatrooms
 - Patienten Communities, Online Gesundheitsplattformen (HealthSuite, MeYou Health, Medelinked)

Die Begriffe “NIS”, “Beobachtungsstudie”, „Big Data“, “RWD” werden häufig synonym verwendet werden

Nutzen von Big Data

- Big Data:
 - im strukturellen Bereich des Gesundheitswesens (Versorgungsforschung)
 - der individuellen Verbesserung des Behandlungsgeschehens (Personalisierte Medizin)
- Elektronische Patientenakte als Schlüsseltechnologie zur Nutzung der Patientendaten innerhalb der GKV.
- In anderen Branchen ist erkennbar, dass die neuen Möglichkeiten der Datenanalytik einen hohen systemischen und persönlichen Nutzen schaffen können.
- Sozialdaten sind ein hohes, schützenswertes Gut sind.

Datenschutzrechtliche Regelungen können verbessert werden

- Befragungsergebnissen der Bitkom:
 - Patienten wollen , ihre Daten zu teilen – selbst höchst persönliche.
 - Kontrolle und einen professionellen, ethischen Umgang
- SHARE = CARE
 - unsere anonymisierten Daten für Analysezwecke zu teilen, dann helfen wir uns und anderen
- **Wir brauchen praktikable Verfahren, für Einwilligungserklärungen von Patienten, um die Datenschätze besser heben zu können.**
 - Datenschutz für die Medizinische Forschung anders organisieren.
 - Patientendaten für Forschungszwecke werden vorher anonymisiert.
 - „Datensparsamkeit“ und „Zweckbindung“ für Forschung und Exploration hinderlich.
 - Löschung der Daten nach 6 Jahren
 - Der Versicherte sollte die Möglichkeit haben seine Gesundheitsdaten selbst zu nutzen bzw. Nutzungsrechte zu übertragen.

