Big Data:
Hype or Help

1. **Definitions and Driving Forces**
2. **Big Data as Prerequisites for Future Medicine**
    1. Conventional Big Data
    2. Unused Big Data
    3. Private Big Data
3. **Conclusion and overall Impact**

Keynote Lecture:
Keynote speakers will describe the political, economic, financial, and / or social issues that will affect the provision of health care in Europe.

Big Data (and its role for medicine):
Hype or Help

Disclaimer:

Pharmacologist, Dean of Med. School, Non-IT Person

Conflict of Interest: **none**

Big Data (and its role for medicine): Hype or Help

Definitions

Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate.
The term often refers simply to the use of predictive analytics or other certain advanced methods to extract value from data, and seldom to a particular size of data set (Wikipedia).

Hype (derived from hyperbole) is promotion, especially promotion consisting of exaggerated claims. (Wikipedia).
A hype tends to disappear after a while (HKK).

Big Data (and its role for medicine): Hype or Help

Big Data is virtually everywhere:

Medicine
Mobility
Consumer relation
Migration
Security
Work Environment
Politics

Big Data (and its role for medicine): Hype or Help

Big Data is virtually everywhere:

Medicine
Mobility
Consumer relation
Migration
Security
Work Environment
Politics

Big Data and Mobility

Mobility is a hallmark of our societies

Based on new sensors and IT, mobility will be redefined.

Serious Consequences, e.g. for ethics:

Partial loss of autonomy (**Disabling Process**)

Partial increase in autonomy (**Enabling Process**)

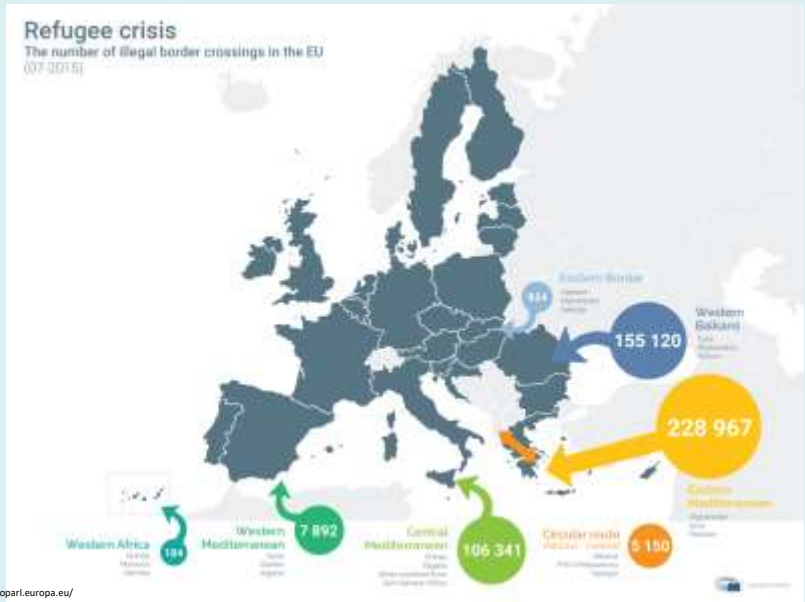
Consumer Relation

The Amazon Example

Based on new sensors and IT, consumer relations will be redefined.

Dash button

Migration



http://www.europarl.europa.eu/

Migration

Refugees on the Greek island of Lesbos take photos of a map
© Iakavos Hartziavrou/AFP/Getty Images

"Our phones and power banks are more important for our journey than anything, even more important than food."
refugeephones.com

Migration

PENN STATE | NEWS

IST researchers explore technology use in Syrian refugee camp

Research shows 86 percent of youth own a mobile handset and more than 50 percent use the internet at least once a day

Stephanie Koons
March 26, 2015

Results of the survey show a high degree of mobile phone and internet use, with 86 percent of youth in their sample owning a mobile handset, and more than half using the internet either once or multiple times per day. There is also a high level of interest in a wide variety of internet based services, particularly social media and news.

The research that Maitland, Xu and their colleagues are conducting is part of an initiative by the Office of the United Nations High Commissioner for Refugees (UNHCR) to collect data on wireless infrastructure and internet use by refugees.

Based on new sensors and IT, migration is enabled

Security

Security by CCTV?

Who evaluates this information?
IT-Supported Algorithms

Security

IT-Supported Algorithms

[...] As Edward Snowden's disclosure of the analysis of bulk data by the US National Security Agency (NSA) and the UK's Government Communications Headquarters (GCHQ) revealed, **the sifting, sorting and triage of vast streams of digital data has become possible because of algorithmic techniques such as pattern recognition, n-gram modelling and distributed querying across cloud databases** [...]. From the real-time stream analysis of online text read by machine learning algorithms to the anomaly-detection algorithms for the discovery of incipient sentiment and human affects, algorithms hold the promise of extending the threshold of human perception and cognition. So, too, do algorithms attend upon, and emerge from, new practices and forms of archival curation, sovereignty, politics and security [...].

Raley R, Amoore L, Security Dialogue, 48(1):1

People fill Via Della Conciliazione boulevard about half a mile away from the facade of St. Peter's Basilica at the Vatican after Pope John Paul II's body was carried across the square into the Basilica for public viewing on April 4, 2005.

Visitors take photos of Pope Francis as he speaks from the central balcony of St. Peter's Basilica at the Vatican, March 13, 2013.

phones, cameras, sharing, covering
→ active part of the communication

washingtonpost.com

Big Data (and its role for medicine): Hype or Help

- Big data is everywhere
- Big data is a Hype
- Big data will not disappear
- Big data does not completely fulfill a hype definition

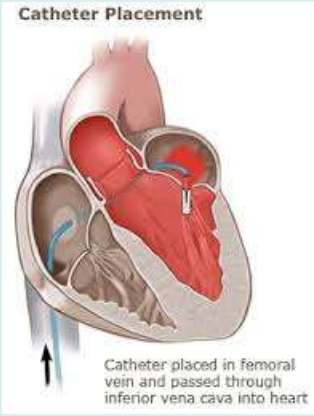
Big Data (and its role for medicine): Hype or Help

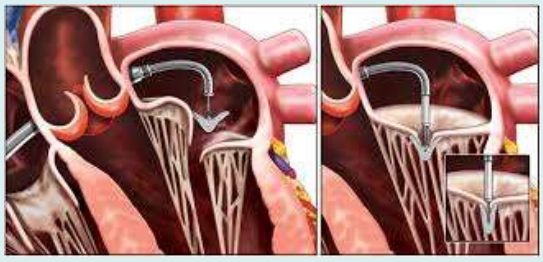
Big Data is virtually everywhere:

Medicine
Mobility
Consumer relation
Migration
Security
Work Environment
Politics

Driving Force for Medicine in 2017: Innovation?

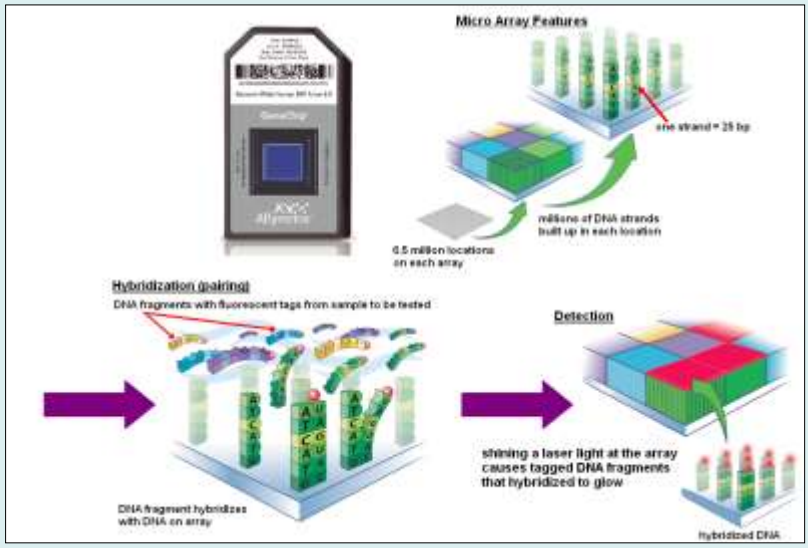
The Mitra Clip Example





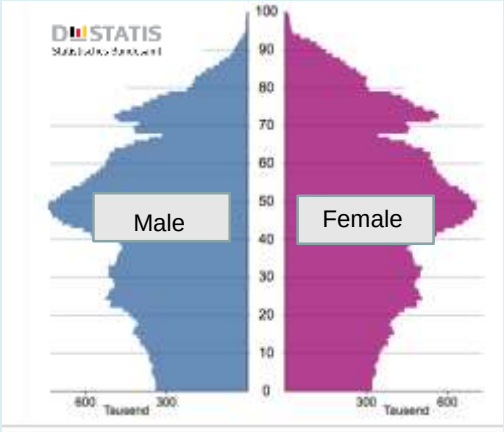
Inserting Cardiac Valves by Catheter
New patient populations
Costs? Decicions?

Driving Force for Medicine: Analytical Technology?



Driving Force for Medicine: Demography

The German Demography 2012



Birth Statistics 1960:
700,000 Male 667,000 Female

-33%

Birth Statistics 1975:
459,000 Male 440,000 Female

Prevalence Mb. Alzheimer

65 to 69 y	1.6 %
80 to 84 y	15.7 %
> 90 y	41 %

3 Mio Alzheimer patients in **2050** out of a total of **70 Mio** (~ 4%)

There will be an enormous pressure from the society towards translational: systems medicine as a solution

Systems Medicine – a Definition

Systems Medicine is the implementation of Systems Biology approaches in medical concepts, research and practice.

Systems medicine involves iterative and reciprocal feedback between experimental and clinical investigations and clinical practice.

Systems Medicine uses computational, statistical and mathematical multiscale analysis and modelling of pathogenetic mechanisms, disease progression and remission, disease spread and cure, treatment responses and adverse events as well as disease prevention both at the epidemiological and individual patient level.

Systems Medicine aims at a measurable improvement of patient health through systems-based approaches and practice.

Modified from: Coordinating Action Systems Medicine

Prerequisites for Systems Medicine

Technology

Big Data

Cohorts / Patients

Big Data in Medicine: hype or help?

Big Data

3 Variants of Big Data:

- **Conventional** Big Data
- **Unused** Big Data
- **Private** Big Data

Big Data in Medicine: hype or help

Conventional Big Data

Data derived from advanced analytical technology (e.g., Omics, Deep Sequencing, Imaging)

Conventional Big Data Example 1
A Clinical Trial (n=1)

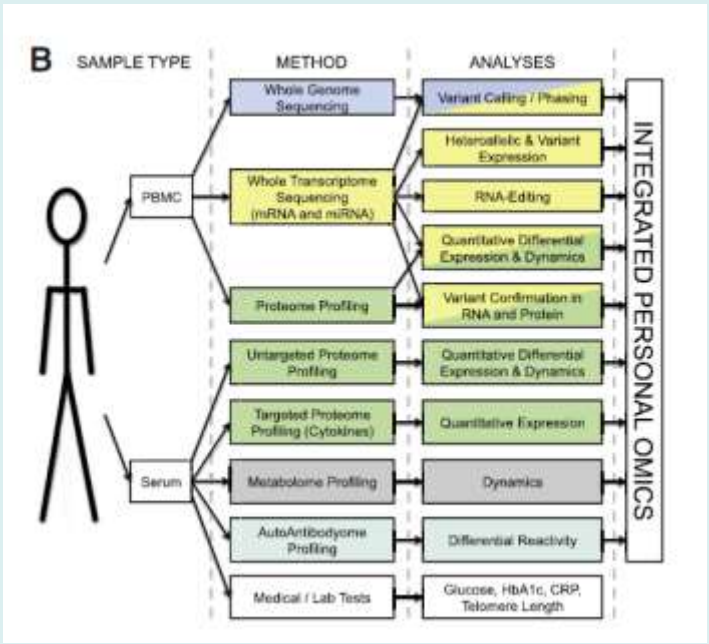
Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes

CELL, March 16, 2012
Chen *et al.*

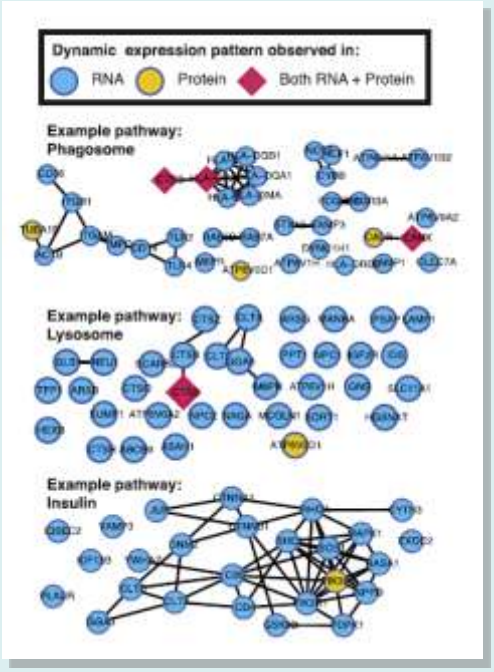
Rui Chen,[1,11] George I. Mias,[1,11] Jennifer Li-Pook-Than,[1,11] Lihua Jiang,[1,11] Hugo Y.K. Lam,[1,12] Rong Chen,[2,12] Elana Miriami,[1] Konrad J. Karczewski,[1] Manoj Hariharan,[1] Frederick E. Dewey,[3] Yong Cheng,[1] Michael J. Clark,[1] Hogune Im,[1] Lukas Habegger,[6,7] Suganthi Balasubramanian,[6,7] Maeve O'Huallachain,[1] Joel T. Dudley,[2] Sara Hillenmeyer,[1] Rajini Haraksingh,[1] Donald Sharon,[1] Ghia Euskirchen,[1] Phil Lacroute,[1] Keith Bettinger,[1] Alan P. Boyle,[1] Maya Kasowski,[1] Fabian Grubert,[1] Scott Seki,[2] Marco Garcia,[2] Michelle Whirl-Carrillo,[1] Mercedes Gallardo,[9,10] Maria A. Blasco,[9] Peter L. Greenberg,[4] Phyllis Snyder,[1] Teri E. Klein,[1] Russ B. Altman,[1,5] Atul J. Butte,[2] Euan A. Ashley,[3] Mark Gerstein,[6,7,8] Kari C. Nadeau,[2] Hua Tang,[1] and Michael Snyder[1,*]

Integrative Personal Omics Profile (iPOP)
from 1 individual over 14 months

Genomic, transcriptomic, proteomic, metabolomic, and Auto-AB profile



Chen *et al.*, Cell, 2012

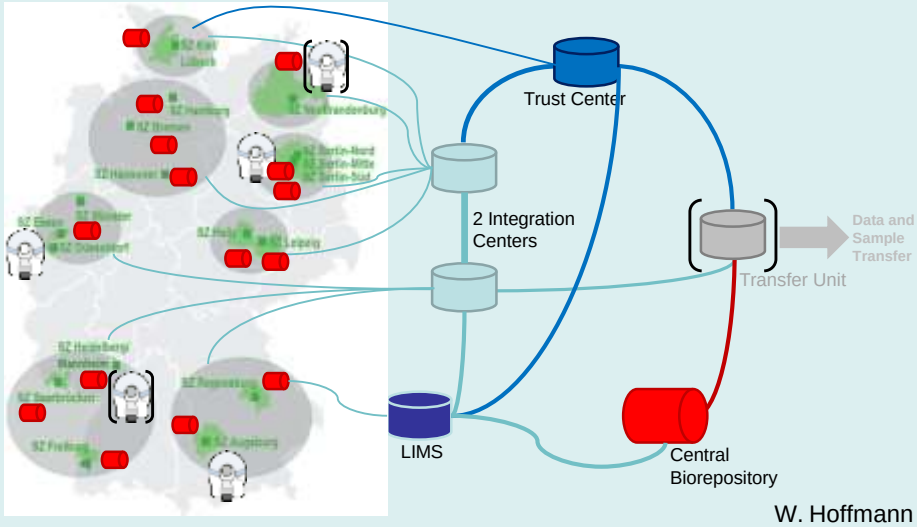


Lessons:

- iPoP is possible
- New insights
- For clinical use: data reduction

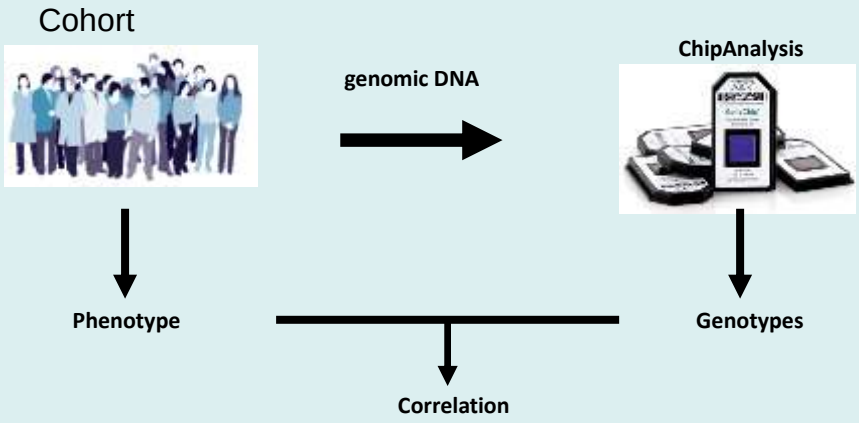
Chen *et al.*, Cell, 2012

Conventional Big Data Example 2
The National Cohort (n = 220,000)



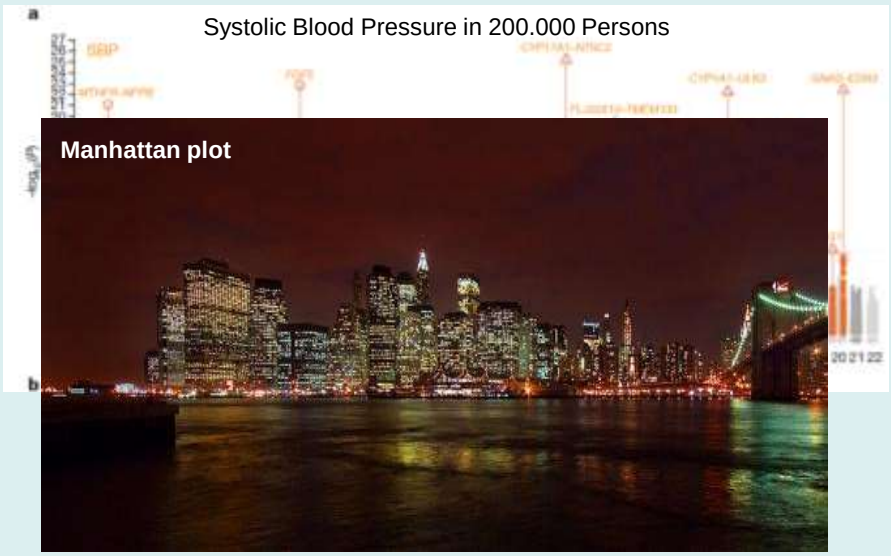
W. Hoffmann

Genome-wide Associations – The Concept



Frequently in combination with big data from several large studies.

Analysis of Genome-wide Associations – Manhattan Plots



Use of Big Data in Medicine

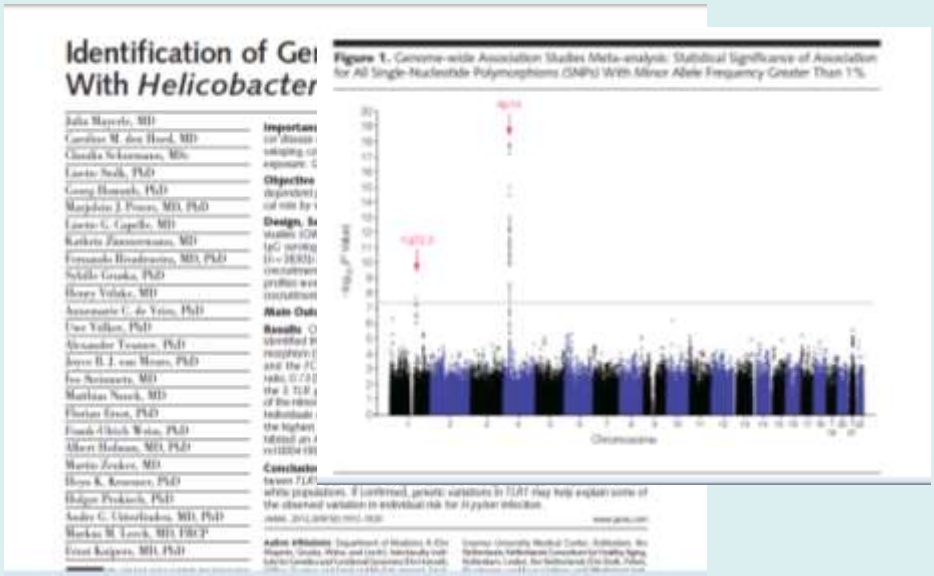
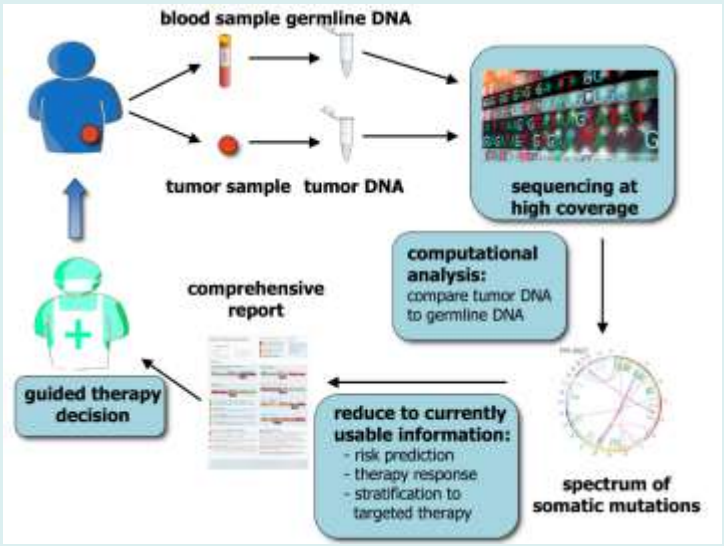

Identification of Ge
With *Helicobacter*

Figure 1. Genome-wide Association Studies Meta-analysis: Statistical Significance of Association for All Single-Nucleotide Polymorphisms (SNPs) With Minor Allele Frequency Greater Than 1%

Big Data -4- NCT Clinical Cancer Program



Massive Genome Sequencing using Illumina HiSeq X Ten

The HiSeq X Ten contains 10 sequencing systems.

HiSeq X™ Ten

Population Power

Composed of 10 HiSeq X Systems, the HiSeq X Ten is the first sequencing platform that breaks the $1000 barrier for a 30x human genome. The HiSeq X Ten System is ideal for population-scale projects focused on the discovery of genotypic variation to understand and improve human health. It can rapidly sequence tens of thousands of samples at high genome coverage, delivering a comprehensive catalog of human variation within and outside coding regions.

- Tens of thousands of whole human genomes per year
- $1000 human genome, including depreciation, sample preparation, and labor

Capacity:	4,500 patients / a	(120x Coverage)
Raw Data:	1,800 TB / a	(5 TB / d)
Total Data including Analysis Data (approx. 2x overhead)	4,000 TB / a	(11 TB / d)
Required growth of storage incl. mirror storage for 2015-2018:	~ 10,000 TB / a	

Big Data @ Heidelberg: 4.500 Patients per Year

Capacity	4.500 Patienten / Jahr (120x Coverage)
Data from Sequencing	1,8 PB / Jahr (5 TB / d)
Total Data including Analytics	4 PB / Jahr (11 TB / d)
Increase in Storage Capacity	ca. 10 PB / Jahr

Big Data (and its role for medicine): Hype or Help

600 Terabytes per day

(Source: Vagata, P., & Wilfong, K. (2014). Scaling the Facebook data warehouse to 300 PB. https://code.facebook.com/posts/229861827208629/)

12 Terabytes per day

(Source: Zhao, L., Sakr, S., Liu, A., & Bouguettaya, A. (2014). Cloud Data Management, Springer)

Sequencing @ DKFZ: 11 Terabytes per day

Data Prerequisites for Systems Medicine

Conventional Big Data

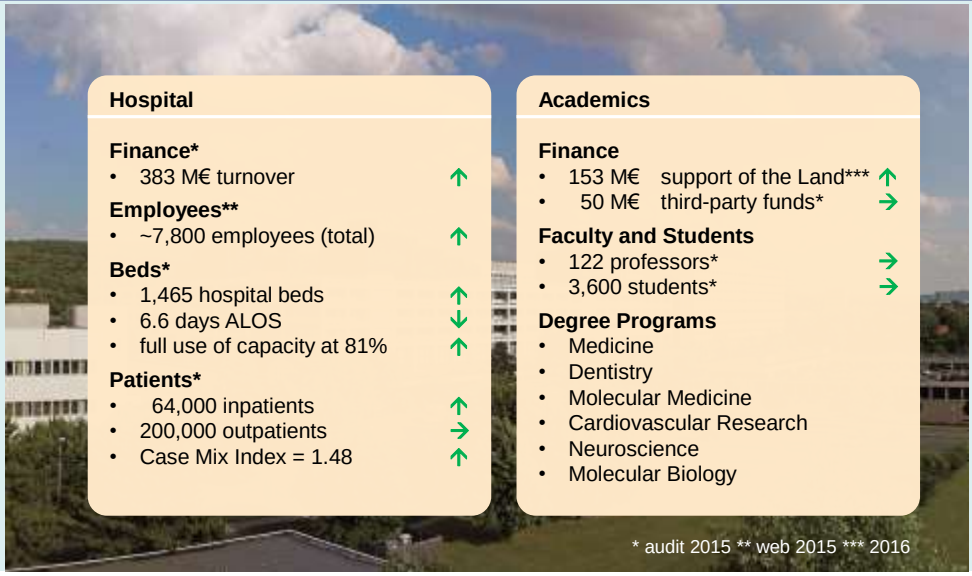
Data derived from advanced analytical technology (e.g., Omics, Deep Sequencing, Imaging)

Unused Big Data

Data derived from standard clinical care

The Academic Medical Center Göttingen

Key Figures



Hospital

Finance*
- 383 M€ turnover ↑

Employees**
- ~7,800 employees (total) ↑

Beds*
- 1,465 hospital beds ↑
- 6.6 days ALOS ↓
- full use of capacity at 81% ↑

Patients*
- 64,000 inpatients ↑
- 200,000 outpatients →
- Case Mix Index = 1.48 ↑

Academics

Finance
- 153 M€ support of the Land*** ↑
- 50 M€ third-party funds* →

Faculty and Students
- 122 professors* →
- 3,600 students* →

Degree Programs
- Medicine
- Dentistry
- Molecular Medicine
- Cardiovascular Research
- Neuroscience
- Molecular Biology

* audit 2015 ** web 2015 *** 2016

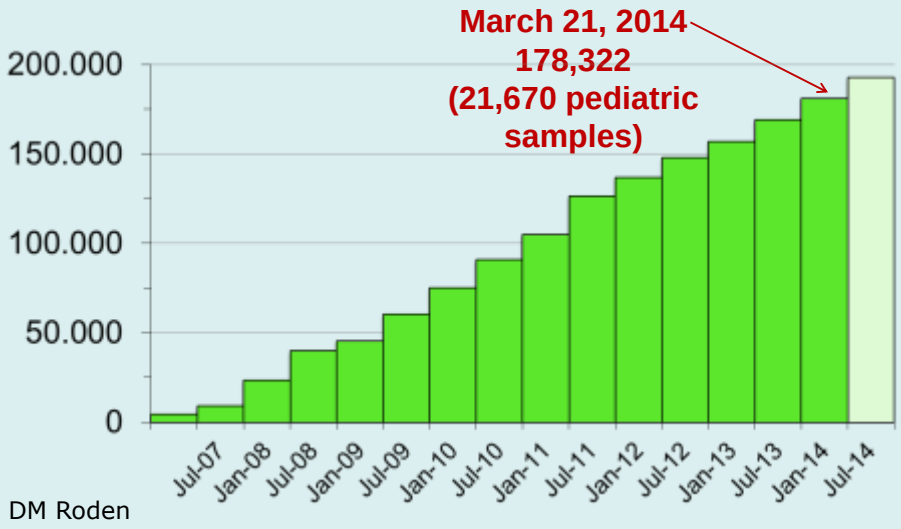
Unused Big Data at the Academic Medical Center Göttingen

0.5 TB per day

Storage currently available	in use	increase	Description and information
1.850 TB			
	260 TB	20 TB	high available - data mirroring in real-time (all medical and administrative systems such as SAP, ICCA, (x)serve)
	140 TB	40 TB	PACS - enlargement in 2015 to 140 TB
	170 TB	20 TB	Snapshots
	120 TB	5 TB	Mailboxes
	200 TB	10 TB	File system (Home VZ, shared drives)
	100 TB		MRT Research area
	20 TB	20 TB	DNA sequences
	20 TB	20 TB	Bio Statistics
	20 TB	20 TB	introduction and operation of emergency department
	10 TB	10 TB	SFB cardiology center
	1.060 TB	165 TB	

BioVU, the Vanderbilt DNA bank

A discovery resource for genomics and pharmacogenomics, linking DNA samples to de-identified electronic medical records



DM Roden

Studies enabled by a very large biobank coupled to Electronic Medical Records

(1) Identifying genomic variants or other markers associated with specific phenotypes:

- Common disease, Rare disease
- Rare outcomes in common disease
- Physiologic traits, lab values: Rare and common
- Drug responses: Rare and common

- Biomarkers for specific disease subtypes: progression, response to drugs
- New drug targets

(2) Identifying phenomic variants associated with specific DNA variants (PheWAS)

(3) Discovery → implementation and outcome assessment

Complex electronic phenotypes can be deployed across multiple EMR systems

Retrograde Translation by Unused Big Data

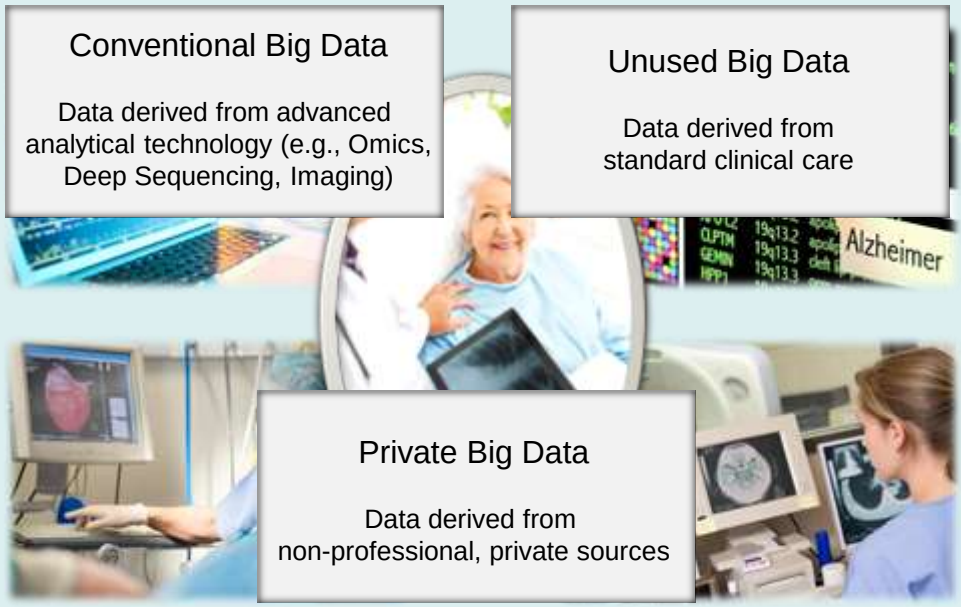
Science ← Translation → Patient Care

PheWAS: Target genotype → association P value

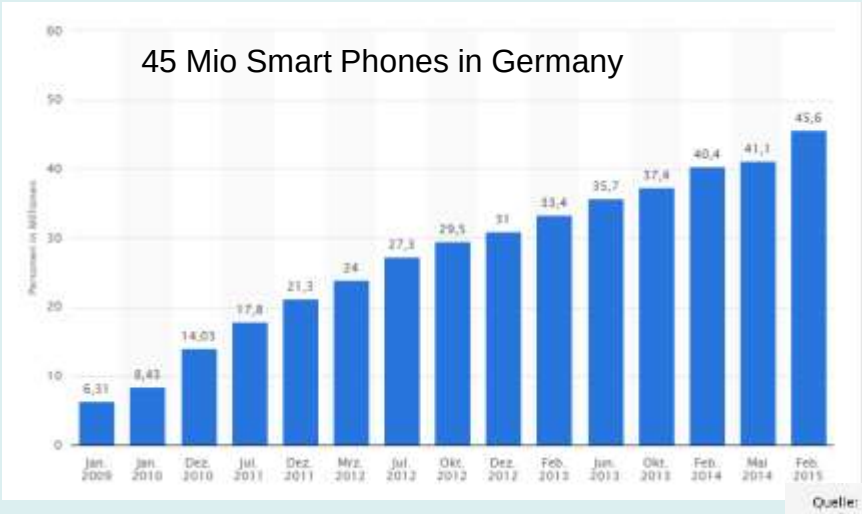
diagnosis code

PheWAS requirement: A large cohort of patients with genotype data and many diagnoses

Data Prerequisites for Systems Medicine



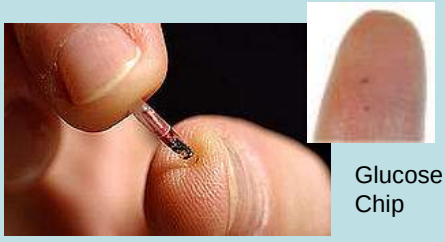
Private Big Data
Smart Phone Users in Germany



Private Big Data (the 24/7 approach)

HR
(RR)

A new type of Systems Medicine



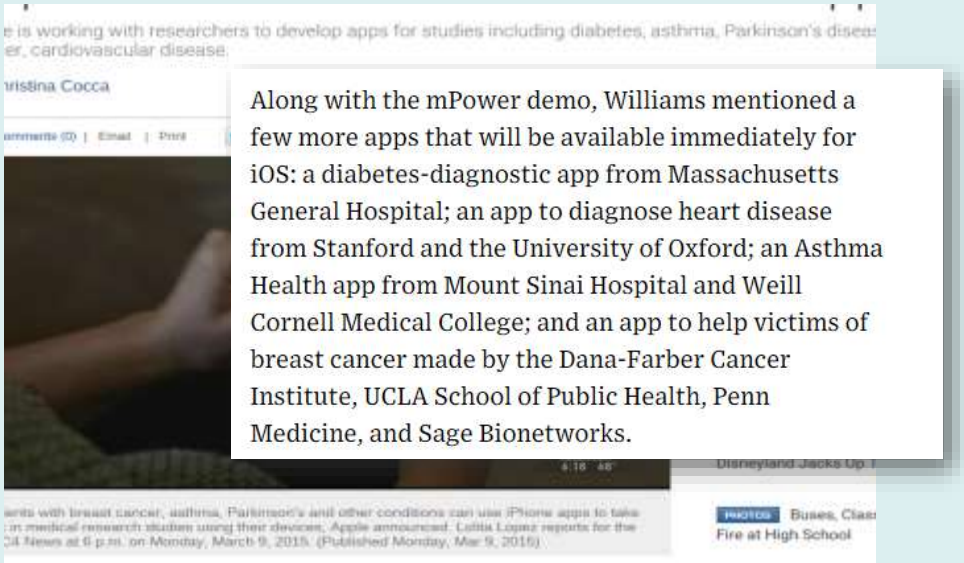
Glucose
Chip

HR
(RR)

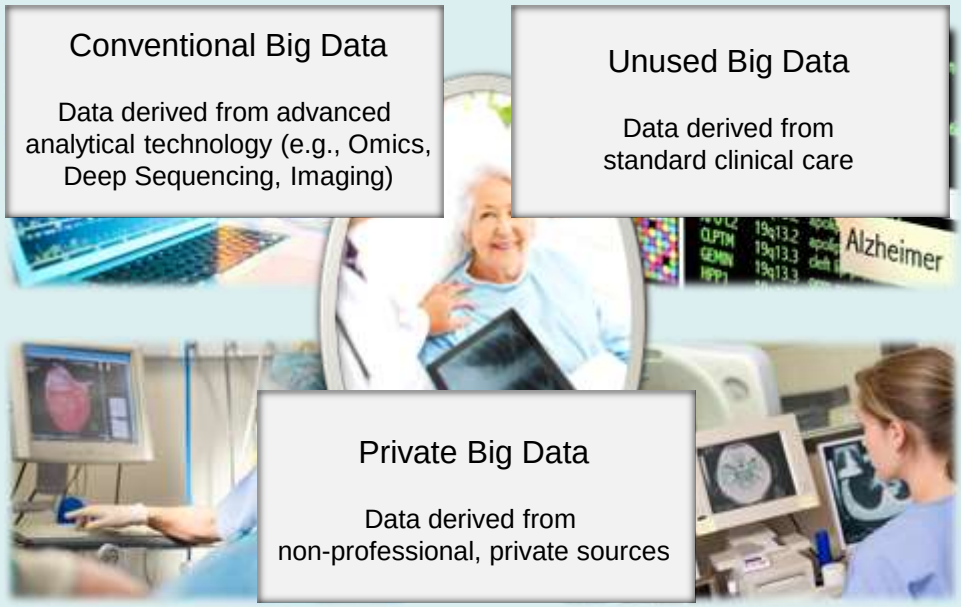
Summary:

- Continuous monitoring of individual data is feasible (eg medication)
- Continuous monitoring of large populations is feasible
- Predefined decision points support care in populations

Systems Medicine based on Private Big Data



Data Prerequisites for Systems Medicine



Prerequisites for Systems Medicine

Big Data

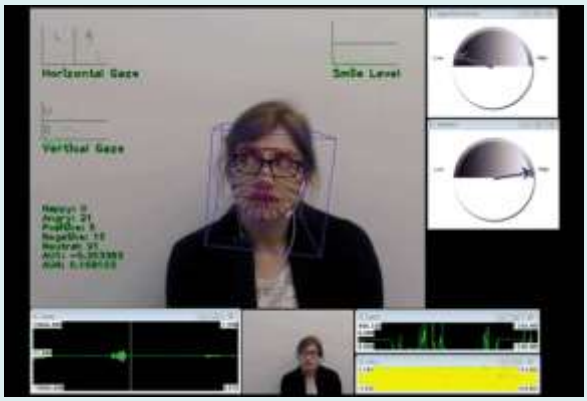
Integrating the 3 Variants of Big Data

- offers an information continuum of the individual
- allows reference to the population even in rare diseases
- enables Systems Medicine (limitations)

Cure in a virtual reality

"Meet Ellie, the machine that can detect depression" – TheGuardian

Her name is Ellie. She introduces herself in a calm voice. Ellie is an avatar, designed to interview mental health patients, gather information about their symptoms and help doctors to develop a diagnosis.



https://mindthehorizon.com/2015/09/21/avatar-virtual-reality-mental-health-tech/

Big Data (and its role for medicine): Hype or Help

IBM Watson =

- 200 million pages of references in 3 seconds
- assessment
- non-structured data

Mayo Clinic and Watson Tackle Clinical Trial Research

Watson at MD Anderson

Oncology Expert Adviser
62.1 Mio $

Big Data Bust: MD Anderson-Watson Project Dies

Obstacles to Big-Data-driven Translation in German Academic Medical Centers:

The largest obstacle for Translational Medicine in German AMCs is the poorly developed information technology.

- No overall concepts at the sites
- No concepts between sites
- Frequently grant associated
- No use of clinical routine data for science

HiGHmed:
Heidelberg
Goettingen
Hannover
Medical Informatics

- Two to three competence centers
- Each center consists of >1 Academic Medical Center
- Alignment of IT-Strategy in a Center
- Translational Approaches as core

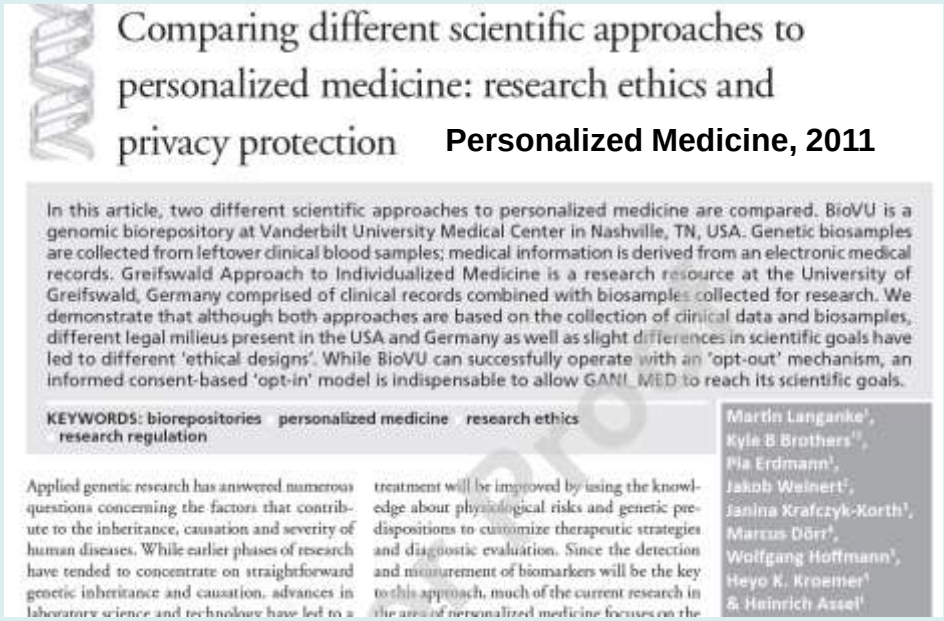
- First Call by the BMBF in Nov. 2015

Non-Medical Questions for the Use of Big Data

Big Data

Legal and Ethical Questions

- Safety
- Privacy
- Ownership
- Participation
- Funding



Comparing different scientific approaches to personalized medicine: research ethics and privacy protection

Personalized Medicine, 2011

In this article, two different scientific approaches to personalized medicine are compared. BioVU is a genomic biorepository at Vanderbilt University Medical Center in Nashville, TN, USA. Genetic biosamples are collected from leftover clinical blood samples; medical information is derived from an electronic medical records. Greifswald Approach to Individualized Medicine is a research resource at the University of Greifswald, Germany comprised of clinical records combined with biosamples collected for research. We demonstrate that although both approaches are based on the collection of clinical data and biosamples, different legal milieus present in the USA and Germany as well as slight differences in scientific goals have led to different 'ethical designs'. While BioVU can successfully operate with an 'opt-out' mechanism, an informed consent-based 'opt-in' model is indispensable to allow GANI_MED to reach its scientific goals.

KEYWORDS: biorepositories personalized medicine research ethics research regulation

Martin Langanke[1], Kyle B Brothers[1,2], Pia Erdmann[1], Jakob Weinert[1], Janina Krafczyk-Korth[1], Marcus Dörr[3], Wolfgang Hoffmann[3], Heyo K. Kroemer[1] & Heinrich Assel[1]

Applied genetic research has answered numerous questions concerning the factors that contribute to the inheritance, causation and severity of human diseases. While earlier phases of research have tended to concentrate on straightforward genetic inheritance and causation, advances in laboratory science and technology have led to a

treatment will be improved by using the knowledge about physiological risks and genetic predispositions to customize therapeutic strategies and diagnostic evaluation. Since the detection and measurement of biomarkers will be the key to this approach, much of the current research in the area of personalized medicine focuses on the

Big Data (and its role for medicine): Hype or Help

A summary:

- Big Data in Medicine is Reality
- We see rapid progress
- The process is a problem

Consequences of Big Data – a wider Frame

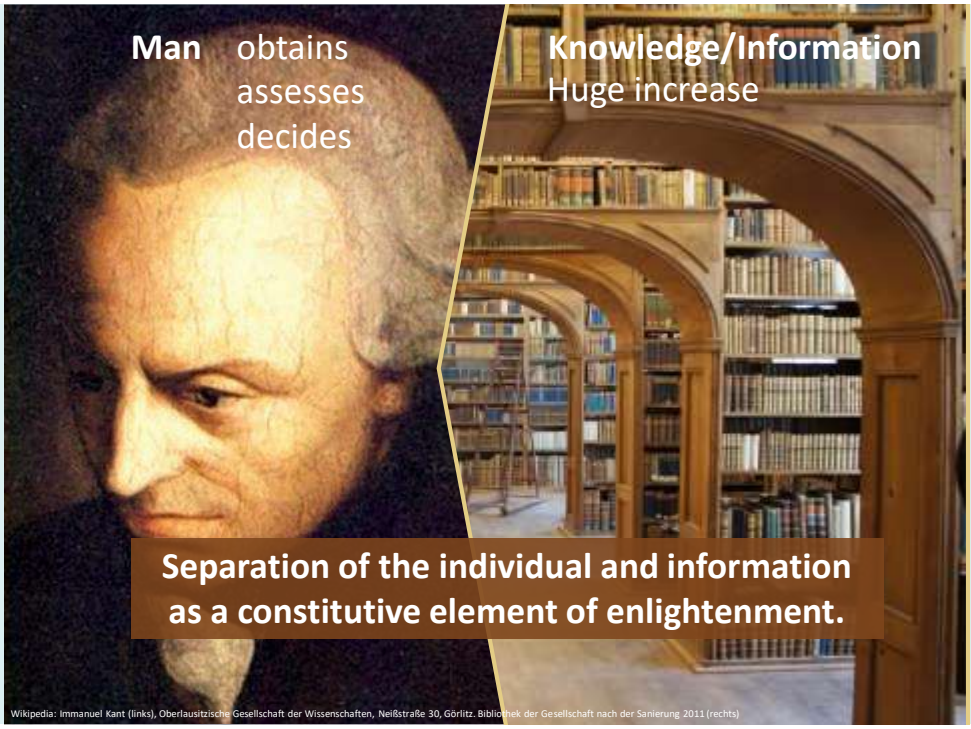
Enlightenment

The **Enlightenment** (or **Age of Reason**) is an era in which cultural and intellectual forces in Western Europe emphasized reason, analysis, and individualism rather than traditional lines of authority.

"Enlightenment is man's emergence from his self-inflicted immaturity.“
Immanuel Kant

Enlightenment is based on information actively used by the individual!

Consequences of Big Data – a wider Frame



Wikipedia: Immanuel Kant (links), Oberlausitzische Gesellschaft der Wissenschaften, Neißstraße 30, Görlitz. Bibliothek der Gesellschaft nach der Sanierung 2011 (rechts)

Big Data and the new Age of Enlightenment (2.0)

- Generation and assessment of information is increasingly autonomous.
- Man is no longer the driving force and will become partially exchangeable.
- Separation between man and information disintegrates (singularity).
- This is an ubiquitous, systemic phenomenon.

Big Data (and its role for Medicine): hype or help?
Three Take Home Messages

- Unique Opportunity based on Big Data
- Responsible Use of Big Data is a Prerequisite
- Foresight Processes need to be established

Acknowledgement:
Information/Figures provided by:

Roland Eils, German Cancer Center, **Heidelberg**

Andre Fischer (DZNE)
Martin Grodzki
Ulrich Sax
Otto Rienhoff, Academic Medical Center, **Göttingen**

Wolfgang Hoffmann
Henry Völzke, Academic Medical Center, **Greifswald**

Dan. M. Roden, Vanderbilt Medical Center, **Nashville**