

Was ist Big Data?

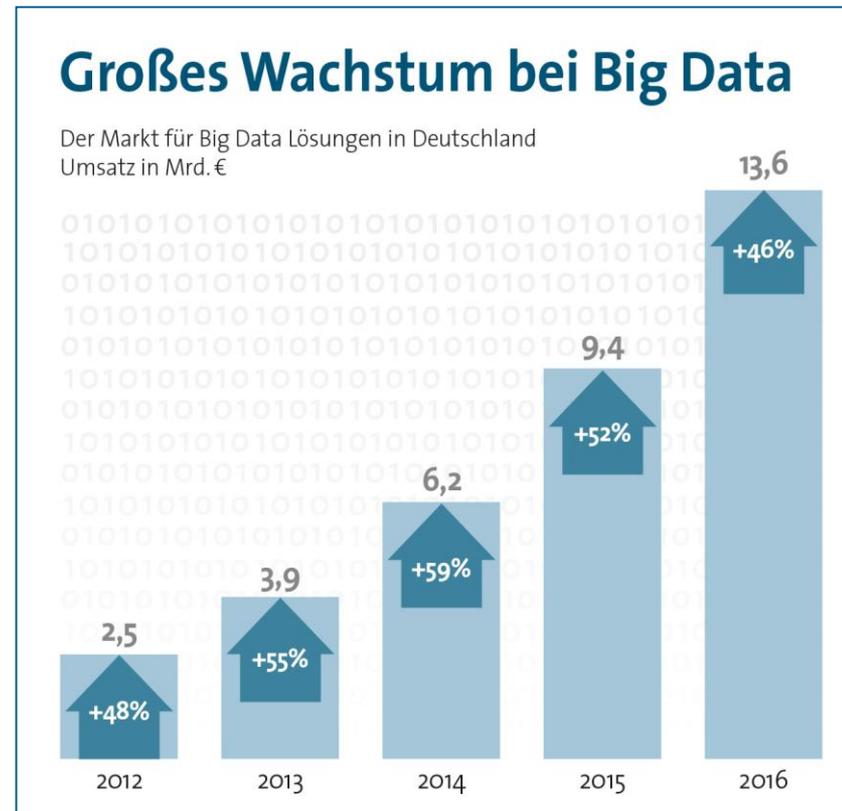
Versuch einer Positionsbestimmung

Wolfgang Marquardt

Vorstandsvorsitzender des Forschungszentrum Jülich
02.06.2015 | Jahrestagung des deutschen Ethikrates

Ganz sicher auch ... ein Hype

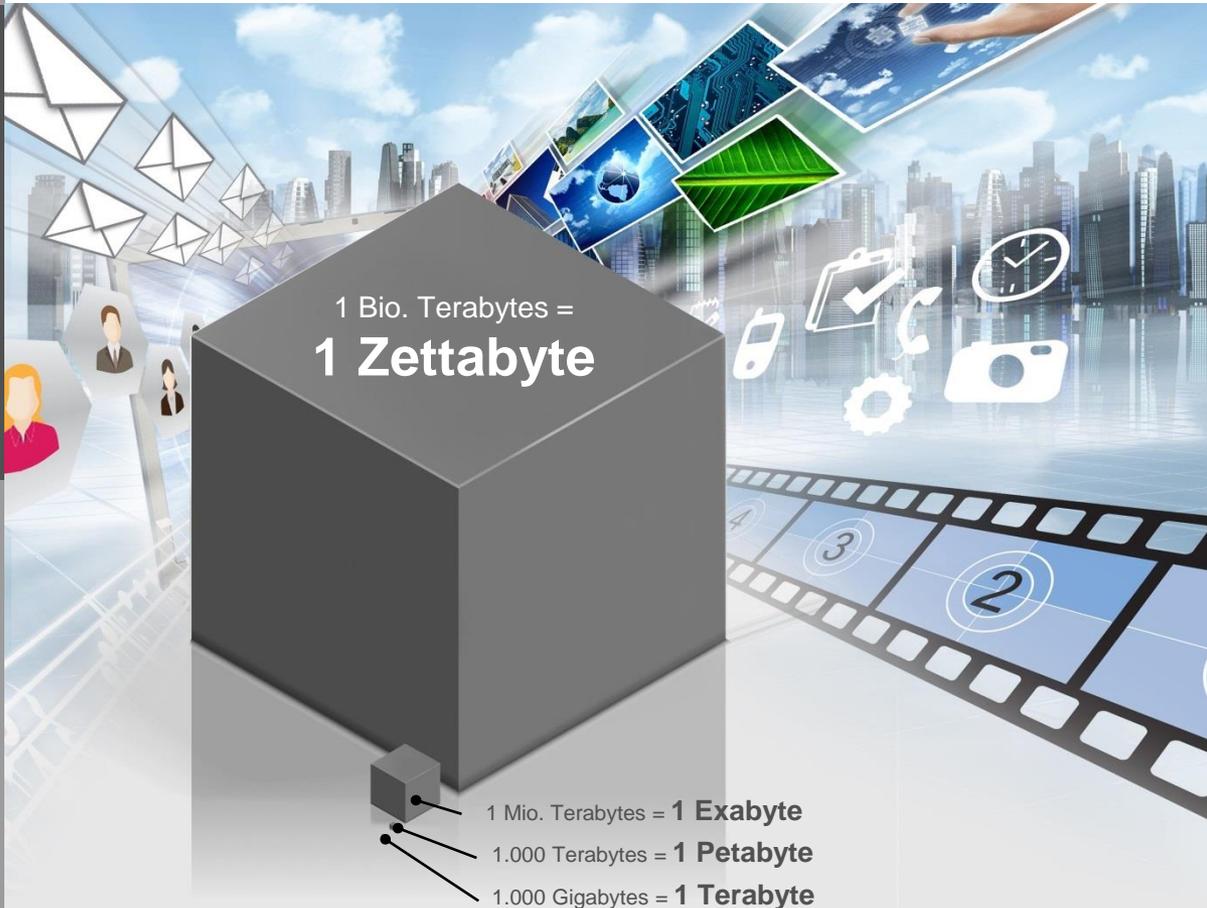
- hohe Erwartungen
- Chancen und Risiken
- Rahmensetzung und Erwartungsmanagement durch die Politik?
- Analyse und Lösungen aus der Wissenschaft?



Quelle: Bitcom, Crisp Research

Big Data

2,7 Zettabyte an Daten 2012 global gespeichert,
1,8 Zettabyte neue Daten in 2015 generiert, ...



+ 20.000 Publikationen
täglich
+ 50 Exabytes neues
„Wissen“ jährlich
= 5 Mio. Libraries of Congress



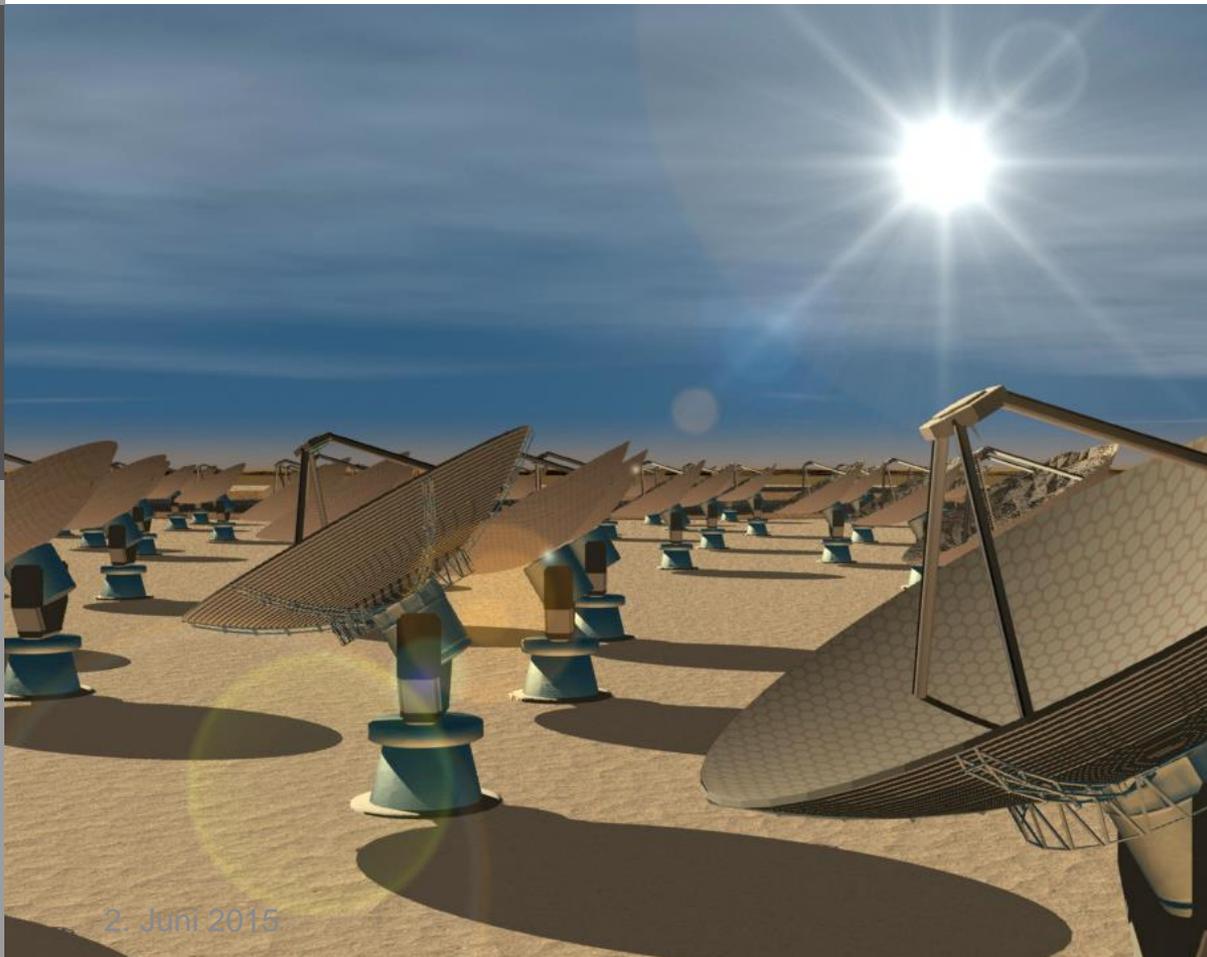
Täglich

+ 600 Terabytes
+ 300 Mio. Fotos
+ 2,7 Mrd. „likes“



The Square Kilometer Array

Das Radioteleskop der nächsten Generation



Verschiebt die Grenzen des beobachtbaren Universums um Milliarden von Galaxien



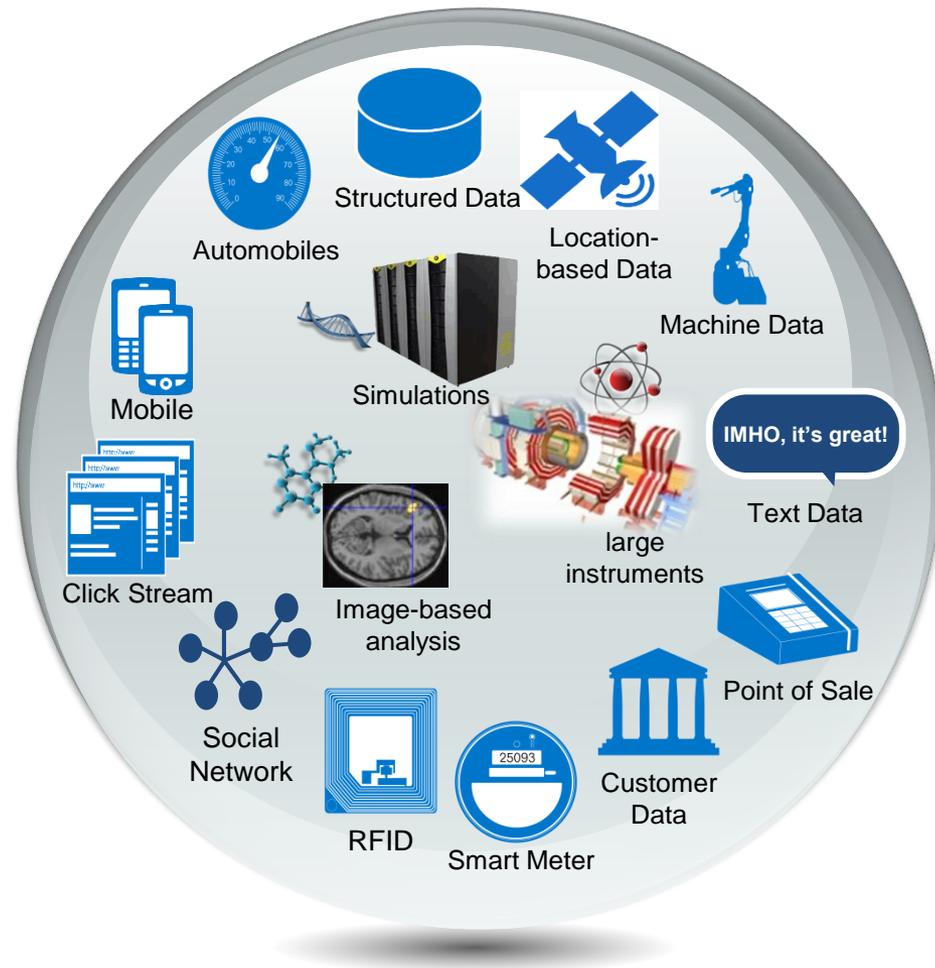
LOFAR-Testfeld in Jülich

Generiert 1 Petabyte alle 20 Sekunden und benötigt Computer, die 1.000-mal schneller sind als der Schnellste der Welt

Big Data – Versuch einer Annäherung

Charakterisierung durch 5 V's

- **Volume**
- **Variety**
- **Velocity**
- **Veracity**
- **Value**

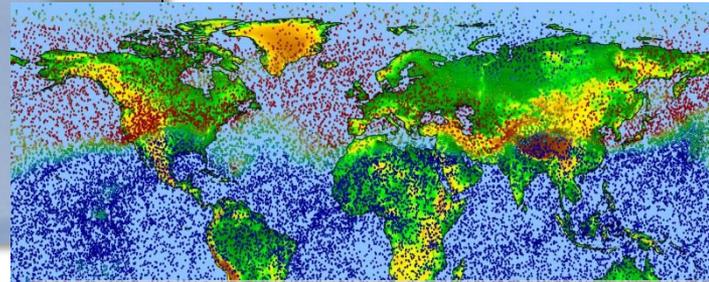


Beiträge von Forschung und Entwicklung

Methoden und Technologien

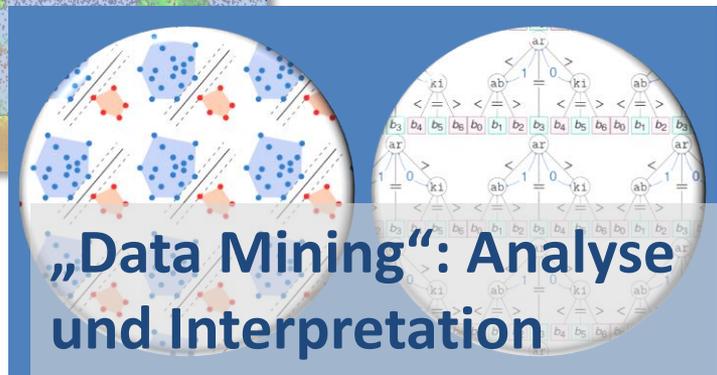


**Datenerfassung,
Speicherung und
Nutzung**



Visualisierung

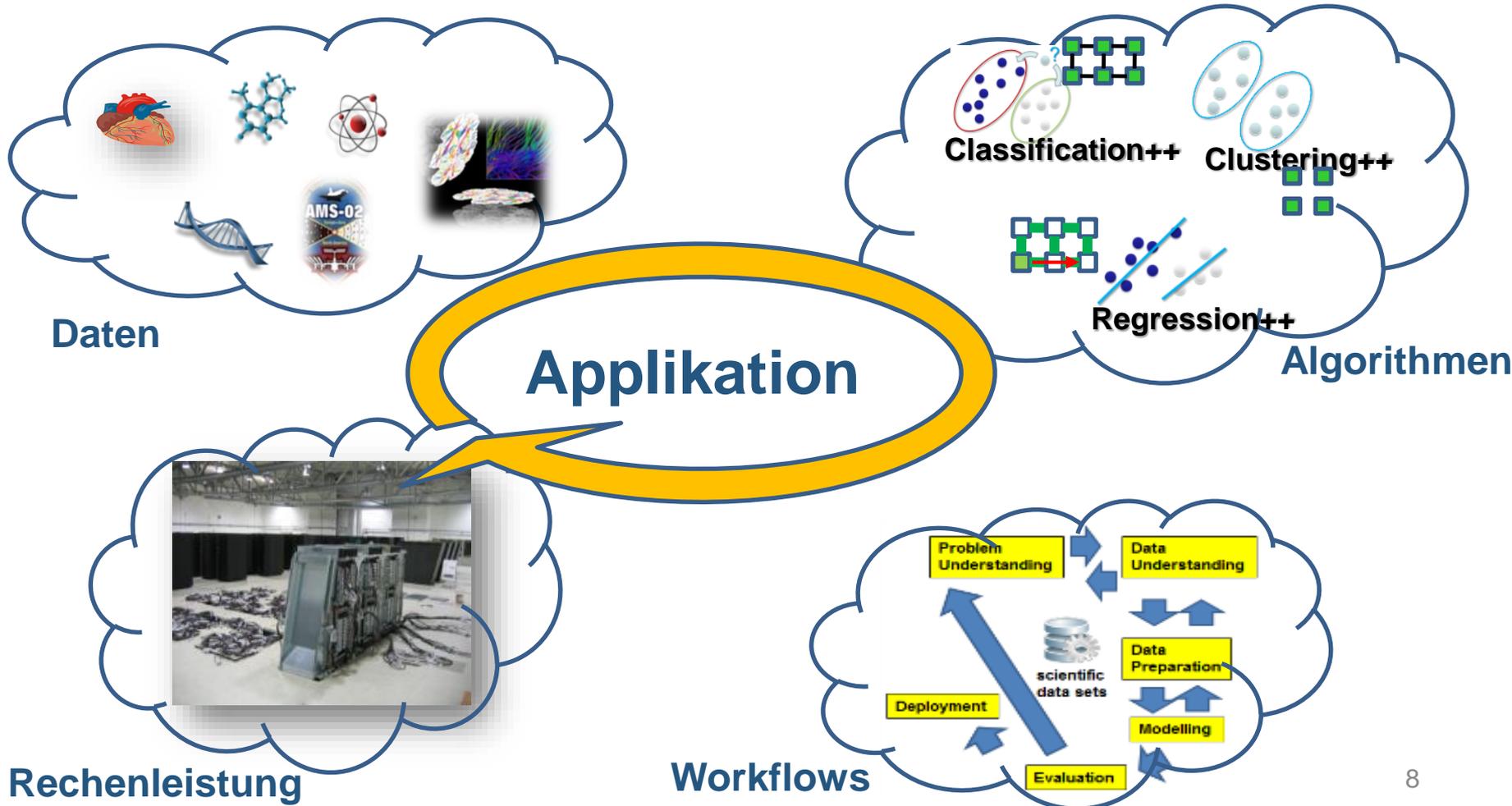
„Data Mining“ macht
„Big Data“ zu „Smart Data“



**„Data Mining“: Analyse
und Interpretation**

Data Mining

Multivariate Regression, Machine Learning, ...



Nationale Kohorte – eine Big Data Challenge

Erfassung, Archivierung, Bereitstellung, Analyse

Datenerhebung

- 200.000 Menschen im Alter von 20-69 Jahren über 10-20 Jahre
- Lebensgewohnheiten, biomed. Marker, ...

Datenanalyse: Korrelation auftretender Krankheiten mit

- genetischen Faktoren,
- Umweltbedingungen,
- sozialem Umfeld
- Lebensstil

Datennutzung

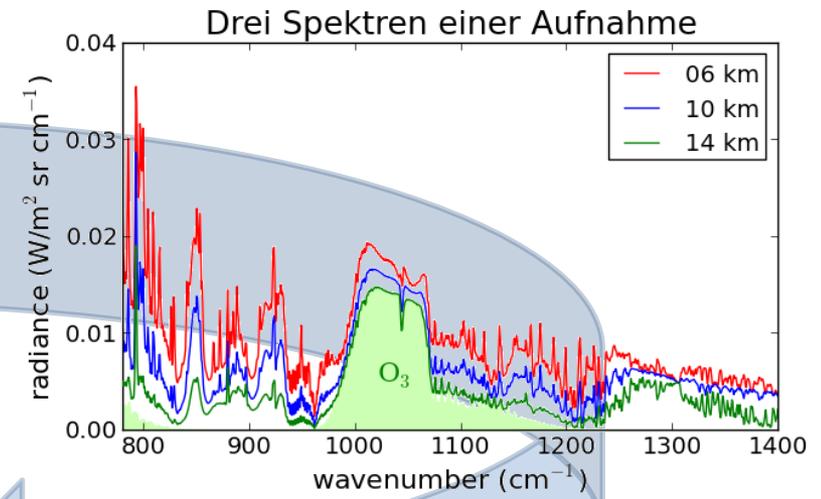
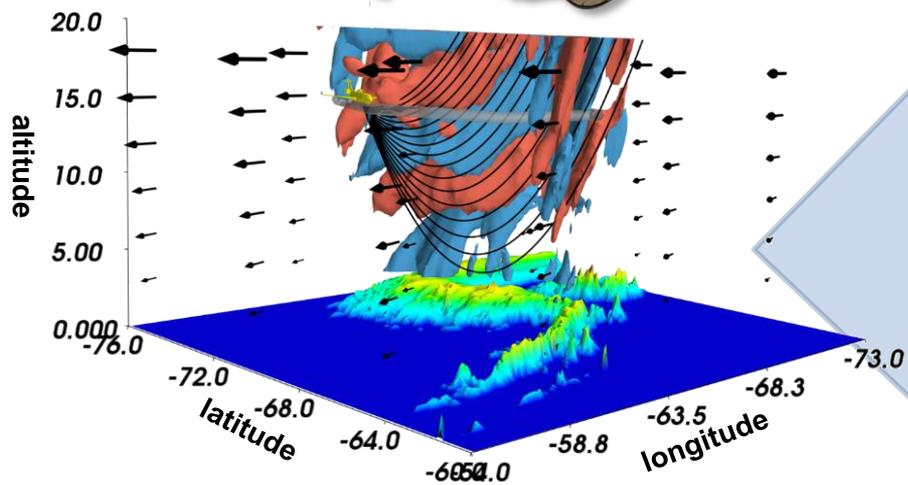
- Ursachen von **Volkskrankheiten** aufklären
- **Risikofaktoren** identifizieren
- wirksame **Prävention**
- **Früherkennung** von Krankheiten



NA
KO
NATIONALE
KOHORTE

Data Mining mit Vorwissen

Verknüpfung von Daten mit Simulationsmodellen



$$\min_x \|\mathbf{F}(\mathbf{x}) - \mathbf{y}\|_2^2 + \lambda \|\Phi(\mathbf{x})\|$$

Simulation und Lösung
des inversen Problems

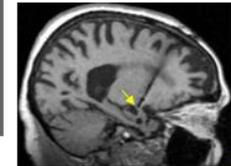
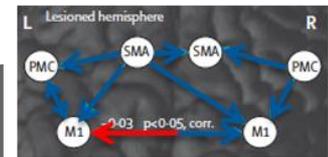
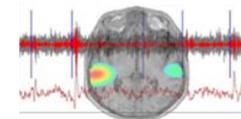
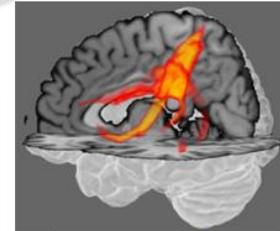
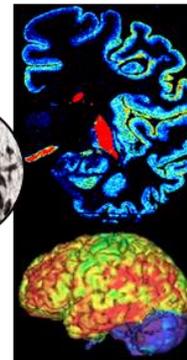
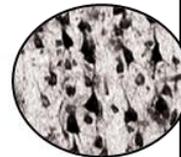
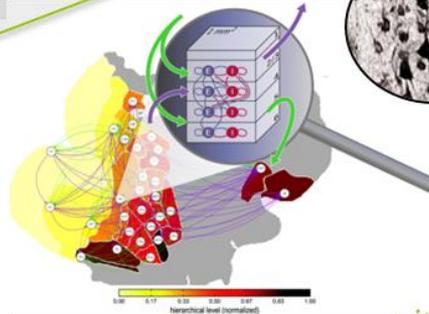
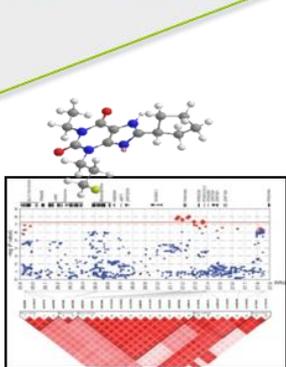
Neue Formen der Wissenschaft: Konvergenz

Aufklärung von Struktur und Funktion des menschlichen Gehirns

Simulation-Lab
Jülich supercomputers
Visualization
BigData
Neuroinformatics



**BIG DATA ANALYTICS
& SIMULATION**



Multiscale in space and time, multimodal

NEUROIMAGING



- 3T MR
- 9.4 MR-PET
- 3T MR-PET
- Electron microscopy
- Image analysis
- 9.4 T animal MR
- Tracer development
- Radiopharmacology
- High-throughput
- microscopical
- imaging

Digitalisierung der Gesellschaft

Produkt- und Prozessinnovationen, neue Geschäftsmodelle ...

- Verkehrsleitsysteme
- Autonome Fahrzeuge
- Gebäudeautomatisierung
- Dezentrale Energieversorgung
- Internet der Dinge: Industrie 4.0
- Analyse und Steuerung des Kaufverhaltens
- Lifestyle-Produkte
- Personalisierte Medizin
- ...



Größer, schneller, besser?

Big Data und Digitalisierung bergen auch Risiken

- Qualität: Daten und Analyse
- Sicherheit: Vulnerabilität von kritischen Infrastrukturen
- Komplexitätsspirale von Angriffs- und Schutzmechanismen
- Datenschutz und Persönlichkeitsrechte
- Gesellschaftliche Akzeptanz erfordert Vertrauen



Qualität: Vorhersage von Grippefällen

„Google Flue Trends“

Data Mining von Web- und Social Media-Daten anstatt ärztliches Berichtswesen in 2700 Praxen mit ca. 30 Mio. Patienten

EPIDEMIOLOGY 14 FEBRUARY 2013 | VOL 494 | NATURE | 155

When Google got flu wrong

US outbreak foxes a leading web-based method for tracking seasonal flu.



Grenzen der Machbarkeit und Vorhersagegüte verstehen!

Datenschutz: Personalisierte Dienstleistungen

Wie persönlich darf es werden?



Öffentliche Wahrnehmung

Karikatur „Die Datenkrake“ – ein Spiegelbild von Ängsten

- Ehrliche Kommunikation von Chancen und Risiken
- Bildung durch Wissenschaft
- Wissens- und Technologietransfer in die Gesellschaft
- Rahmensetzung durch die Politik



Dapp (2014)

Big Data und Digitalisierung werden die Welt verändern!

**Politik, Wissenschaft und Wirtschaft
müssen gemeinsam
Verantwortung für die Ausgestaltung
übernehmen.**