

# Big Data in Medicine: Hype or Hope for the Patient?

Prof. Dr. Michael Krawczak

Institute of Medical Informatics and Statistics

Kiel University, Germany

TMF e.V., Berlin, Germany

Bone Innovation Summit 2019

Lübeck, 13-14 February 2019

# Paradigms of Future Medicine

## Translational Research

“In short, [translational research] is the process of turning appropriate biological discoveries into drugs and medical devices that can be used in the treatment of patients.”

*wikipedia.org*

## Precision Medicine

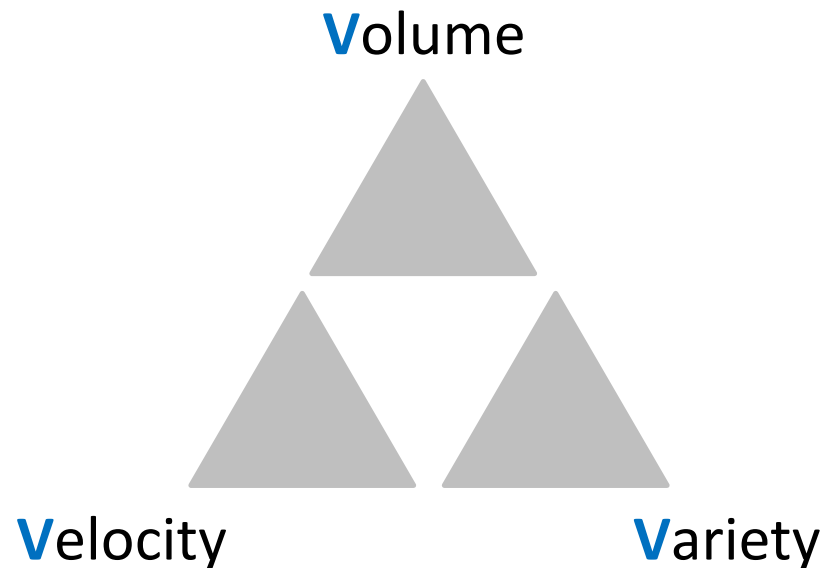
“Precision medicine is an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person.”

*www.nih.gov*

# What is "Big Data"?

“Big Data is **high-volume**, **high-velocity** and/or **high-variety** information assets that demand cost-effective, innovative forms of information processing that enable **enhanced insight, decision making, and process automation.**”

*Gartner IT Glossary*



## Angry white guys

Rafael Behr on the culture wars consuming Cameron's Conservatives

Laurie Penny  
Remembering  
Emily Davison,  
the suffragette hero

Ed Smith  
In Britain, as in  
cricket, the north  
south divide

# New Statesman

Free thinking since 1913

24-30 May 2013/£3.50

## How Big Data took over our lives

By Steven Poole

facebook knows what you like

Google knows what you want

Apple knows what you buy

amazon knows what you read

Microsoft knows where you live

PLUS Rhiannon Lucy Cosslett When should a woman have a baby?



## Statistical and Machine-Learning Data Mining

Techniques for Better Predictive Modeling and Analysis of Big Data

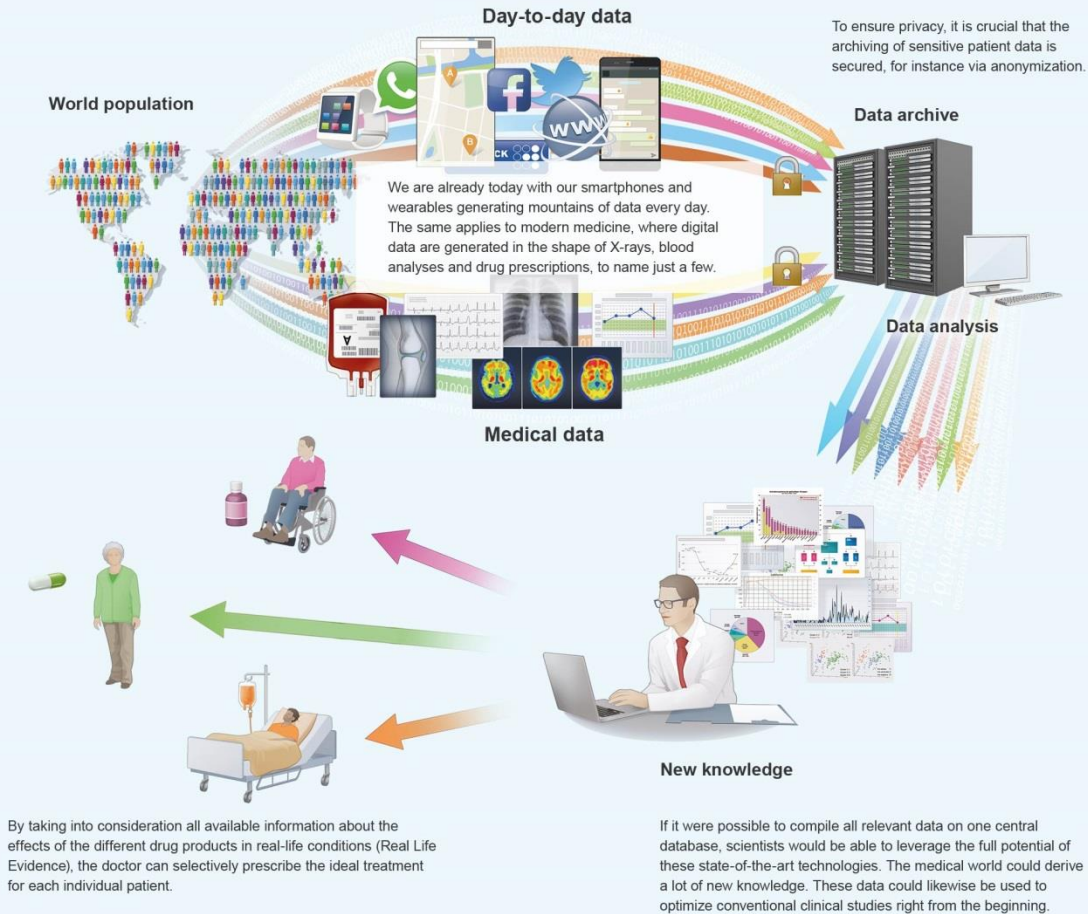
Second Edition



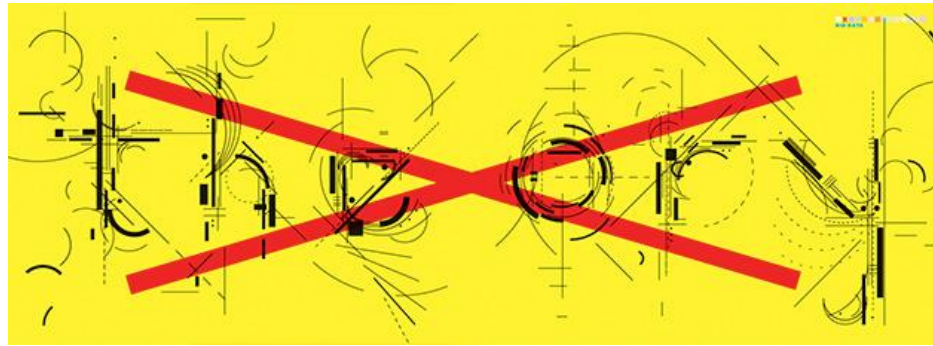
Bruce Ratner

CRC Press  
Taylor & Francis Group

## Big data in medicine



# Future of Science ?



© Wired Magazine

“**Correlation supersedes causation**, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all.”

*Chris Anderson, Chief Editor, Wired Magazine*

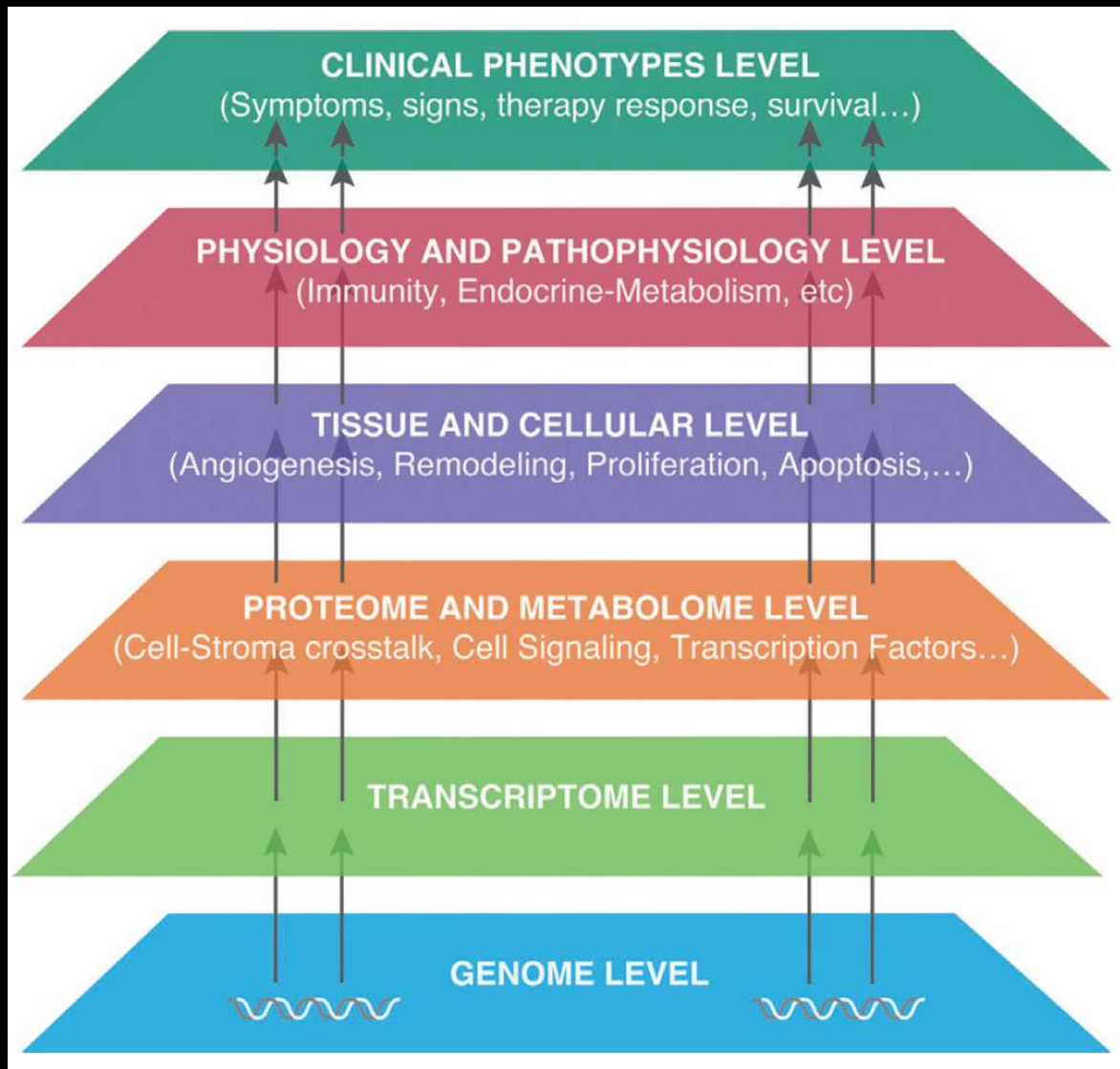
# Big Data in Medicine

“Finding **new correlations** is expected to greatly contribute to the development of medicine [but] correlations suggested by the analysis of big data should be **validated through clinical studies.**”

*Kanda E (2015) Use of Big Data in Medicine, Ren Repl Ther 1:4*

“Big data comes into play around **aggregating more and more information** around multiple scales for what constitutes a disease.”

*Eric Schadt  
Icahn Institute for Genomics and Multiscale Biology  
Mount Sinai Hospital, New York, USA*



*Blanco-Gomez A, et al. (2016) Bioessays 38: 664-673.*



# Das Gen-Orakel

**Medizin** Jahrelang suchten Forscher im Erbgut vergebens nach den Ursachen der großen Volkskrankheiten. Nun endlich melden sie Erfolge: Dank Auswertung Hunderttausender DNA-Tests gelingt es immer besser, das gesundheitliche Schicksal von Menschen vorherzusagen.

**D**ieser Mann war anders als andere Männer. Nie zuvor hatte John Kauwe einen Menschen gesehen, der selbst in den großzügig bemessenen Sitzen der First Class seine Beine so jämmerlich zusammenfallen musste. Kauwe kam mit dem Riesen neben ihm ins Gespräch. Er hatte Shawn Bradley vor sich, der vor rund 20 Jahren als einer der größten Spieler der Basketballliga NBA für Furor gesorgt hatte. Bradley wiederum war erfreut, als sich sein Sitznachbar als Genetiker vorstellte. Ob Kauwe denn das Geheimnis seines Riesenwuchses (2,29 Meter) lüften könne, fragte er.

Aus der Flugzeugbekanntschaft wurde ein Forschungsprojekt. Inzwischen hat Kauwe veröffentlicht, was dabei herausgekommen ist: Das vermeintliche Geheimnis gibt es gar nicht. In Bradleys Erbgut findet sich keine Spur eines Goliathgens. Seine Erbanlagen sind normal, außergewöhnlich ist nur deren Mischung. Kauwe und seine Kollegen wiesen viele Hundert genetische Varianten nach, von denen jede Bradley ein klein wenig weiter aufschießen ließ. Zusammengekommen machten sie ihn übergroß.

Botschaften wie jene, die der Genforscher Kauwe im Erbgut des Basketballers fand, begegnen seinen Kollegen derzeit allerorten: Sie stellen fest, dass es nicht einzelne Genvarianten gibt, die uns aggressiv oder trübsinnig machen. Die viel beschworenen Intelligenz- oder Schwulengene: nichts als Phantome. Und auch für Herzinfarkt, Demenz oder Schizophrenie sind nicht bestimmte Erbanlagen verantwortlich.

Die Erklärung: Gene wirken nicht einzeln, sie stehen in komplexem Wechselspiel miteinander. Jedes unserer Merkmale wird von Hunderten oder gar Tausenden Varianten im Erbgut bestimmt, und umgekehrt hat jedes Gen seinen Anteil an Hunderten unserer Eigenschaften.

Doch so verworren dieses Dickicht auch scheinen mag: Inzwischen beginnen Forscher vorzudringen ins genetische Wirkungsgeflecht. Zwar verstehen sie die Rolle all der zahllosen Gene und ihrer Varianten nicht, dazu sind deren Regelkreise viel zu verwirrend. Doch auch ohne ein wirkliches Verständnis gelingt es mit Com-

puterhilfe, statistische Wahrscheinlichkeitsaussagen zu treffen, wie groß, wie klug oder wie gesund der Träger eines bestimmten Genoms vermutlich sein mag.

Sekar Kathiresan hat die Kunst der genetischen Kaffeesatzleserei weit vorangetrieben. Der Kardiologe vom Massachusetts General Hospital hat in »Nature Genetics« sogenannte polygene Risikowerte für fünf häufige Krankheiten vorgestellt. Schon im Kindesalter kann er demnach am Erbgut ablesen, wie wahrscheinlich es ist, dass ein Mensch dereinst einmal an Diabetes Typ 2, an Arteriosklerose der Herzkranzgefäße, an Brustkrebs, chronisch entzündlicher Darmerkrankung oder an Vorhofflimmern leiden wird.

»Bald wird der Begriff »polygenes Risiko« so geläufig sein wie heute der Begriff »Cholesterin«, prophezeit Kathiresan. In der Medizin der Zukunft werde die Prä-

## Die genetische Wirklichkeit hat mit den mendelschen Regeln wenig zu tun.

vention im Mittelpunkt stehen, das Errechnen von Krankheitsrisiken bekomme da zentrale Bedeutung.

In einer Klinik für präventive Herzmedizin behandelt Kathiresan Menschen, die, oft geschockt vom unvermittelten Herztod naher Verwandter, nach Erklärungen suchen für den Schicksalsschlag. »Bisher beruht unsere Anamnese auf drei Säulen«, sagt Kathiresan. »Erstens: Wir befragen die Leute nach ihren Gewohnheiten – also ob sie rauchen, was sie essen oder wie viel sie sich bewegen. Zweitens: Wir bestimmen ihre Blutfettwerte. Und drittens: Wir machen klinische Untersuchungen, mit EKG zum Beispiel, mit Angiografie oder mit Ultraschall.« Nun aber sei die Zeit reif, eine vierte Säule ins Gebäude der Medizin einzuziehen: die polygenen Risikowerte.

Ein Abgleich mit Patientendaten in zwei Bostoner Krankenhäusern gibt ein Gefühl für den Vorhersagewert: 20 000 Krankenakten durchforsteten Kathiresan

und seine Kollegen. Sie stellten dabei fest: Bei denjenigen Patienten, die einen Herzinfarkt erlitten hatten, war der polygene Risikowert im Vergleich zu anderen Patienten im Schnitt tatsächlich auf das Vierfache erhöht.

Es setzt sich damit jetzt eine Form der Genetik durch, die sich maßgeblich unterscheidet von dem, was in den Schulen gelehrt wird. Seit vielen Jahrzehnten lernen die Schüler dort, mit dominanten und rezessiven Merkmalen zu hantieren und mit Stammbäumen, die aus Quadraten und Kreisen bestehen.

Als im Jahr 2000 die Entzifferung des menschlichen Genoms verkündet wurde, da schien es, als erlebten diese mendelschen Vererbungsregeln ihren ultimativen Triumph. Die Grundlage für ein genetisches Verständnis des Menschen sei gelegt; nun werde es nur noch 10, höchstens 15 Jahre dauern, bis der Schlüssel zur Behandlung großer Volksleiden wie Alzheimer und Diabetes gefunden sei, versprach Francis Collins, seinerzeit Leiter des »Human Genome Project«. Bei Diabetes Typ 2 zum Beispiel vermutete er, dass nur rund ein Dutzend Gene am Werk seien, die man nun im Handumdrehen dingfest machen werde.

Zehn Jahre später war Ernüchterung eingetreten. Der medizinische Nutzen des Genomprojekts sei »fast gleich null«, sagte im SPIEGEL-Gespräch (26/2010) der nie um drastische Formulierungen verlegene Genforscher Craig Venter, der selbst eine der Schlüsselfiguren der Genomentzifferung gewesen war. »Jedermann hat erwartet, im Erbgut wundersame Ja-Nein-Antworten zu finden: »Ja, Sie werden Krebs kriegen« oder »Nein, Sie werden keinen Krebs kriegen«. Aber so einfach ist es eben nicht«, sagte er.

Denn die genetische Wirklichkeit hat mit den mendelschen Regeln wenig zu tun. »Die sind in den meisten Fällen bloße Labor-Artefakte«, konstatiert der Genforscher Steven McCarroll von der Harvard-Universität. »Aber Mendel steckt eben tief in den Köpfen.«

Sehr viele Erkenntnisse seines Fachs beruhten auf Experimenten mit Labormäusen, erklärt McCarroll. Die aber seien durch langjährige Inzucht genetisch bei-



Wikimedia Commons

*The Fall of Icarus*  
Jacob Peter Gowy (1615-1661)

# Medical Research Data Management

## Status quo

- fragmented and inadequate research data management
- lack of institutionalized sustainability
- lack of cross-institutional transparency
- poor accessibility and usability of data from health care context
- lack of interoperability of data resources
- no common identity management



*"A journey of a thousand miles begins with a single step."  
Laozi (604-531 B.C.)*

## Die Fakten

### Wer wird gefördert?

Konsortien, denen mindestens zwei Universitätskliniken sowie ggf. weitere Partner (z. B. Forschungseinrichtungen, Hochschulen, Unternehmen, private Kliniken oder andere Träger der Gesundheitsversorgung) angehören

### Was wird gefördert?

Konzeptphase (9 Monate):

Ausarbeitung von Konzepten für

- Datenaustausch und Datennutzung
- den Aufbau von Datenintegrationszentren
- konkrete Anwendungsfälle von nutzbringenden IT-Lösungen
- die zukunftsgerichtete Aufstellung der Medizin-informatik in Forschung, Lehre und Weiterbildung

Aufbau- und Vernetzungsphase (4 Jahre):

- Aufbau von Datenintegrationszentren
- Entwicklung und Implementierung von IT-Lösungen für eine oder mehrere konkrete Anwendungen
- Förderung von Nachwuchsgruppen

### Frist

31. März 2016: Einreichung von Anträgen für die Konzeptphase

### Informationsveranstaltung

27. November 2015 im BMBF in Berlin

### Weitere Informationen

→ [www.gesundheitsforschung-bmbf.de/de/medizininformatik.php](http://www.gesundheitsforschung-bmbf.de/de/medizininformatik.php)



### Kontakt

DLR Projektträger  
Gesundheitsforschung  
Heinrich-Konen-Straße 1  
53227 Bonn  
Tel.: 0228-3821-1210  
Fax: 0228-3821-1257  
[www.gesundheitsforschung-bmbf.de](http://www.gesundheitsforschung-bmbf.de)

## Impressum

### Herausgeber

Bundesministerium für Bildung und Forschung (BMBF)  
Referat Methoden- und Strukturentwicklung  
in den Lebenswissenschaften  
11055 Berlin

### Bestellungen

Publikationsversand der Bundesregierung  
Postfach 48 10 09, 18132 Rostock  
E-Mail: [publikationen@bundesregierung.de](mailto:publikationen@bundesregierung.de)  
Internet: <http://www.bmbf.de>  
oder per Tel.: 030 18 272 272 1  
Fax: 030 18 10 272 272 1

### Stand

November 2015

### Druck

W. Bertelsmann Verlag, Bielefeld

### Gestaltung

W. Bertelsmann Verlag, Bielefeld; Christiane Zay

### Bildnachweis

außen: iStock, Squaredpixels; Presse- und Informationsamt der Bundesregierung, Steffen Kugler; (Porträt Prof. Dr. Johanna Wanka);  
innen: Getty, Ariel Skelley; shutterstock, sfam\_photo

Dieser Flyer ist Teil der Öffentlichkeitsarbeit des Bundesministeriums für Bildung und Forschung; er wird kostenlos abgegeben und ist nicht zum Verkauf bestimmt.

[www.bmbf.de](http://www.bmbf.de)



Bundesministerium  
für Bildung  
und Forschung

## Förderkonzept Medizininformatik

Daten vernetzen – Gesundheitsversorgung verbessern



# BMBF Funding Scheme “Medical Informatics”

SPONSORED BY THE

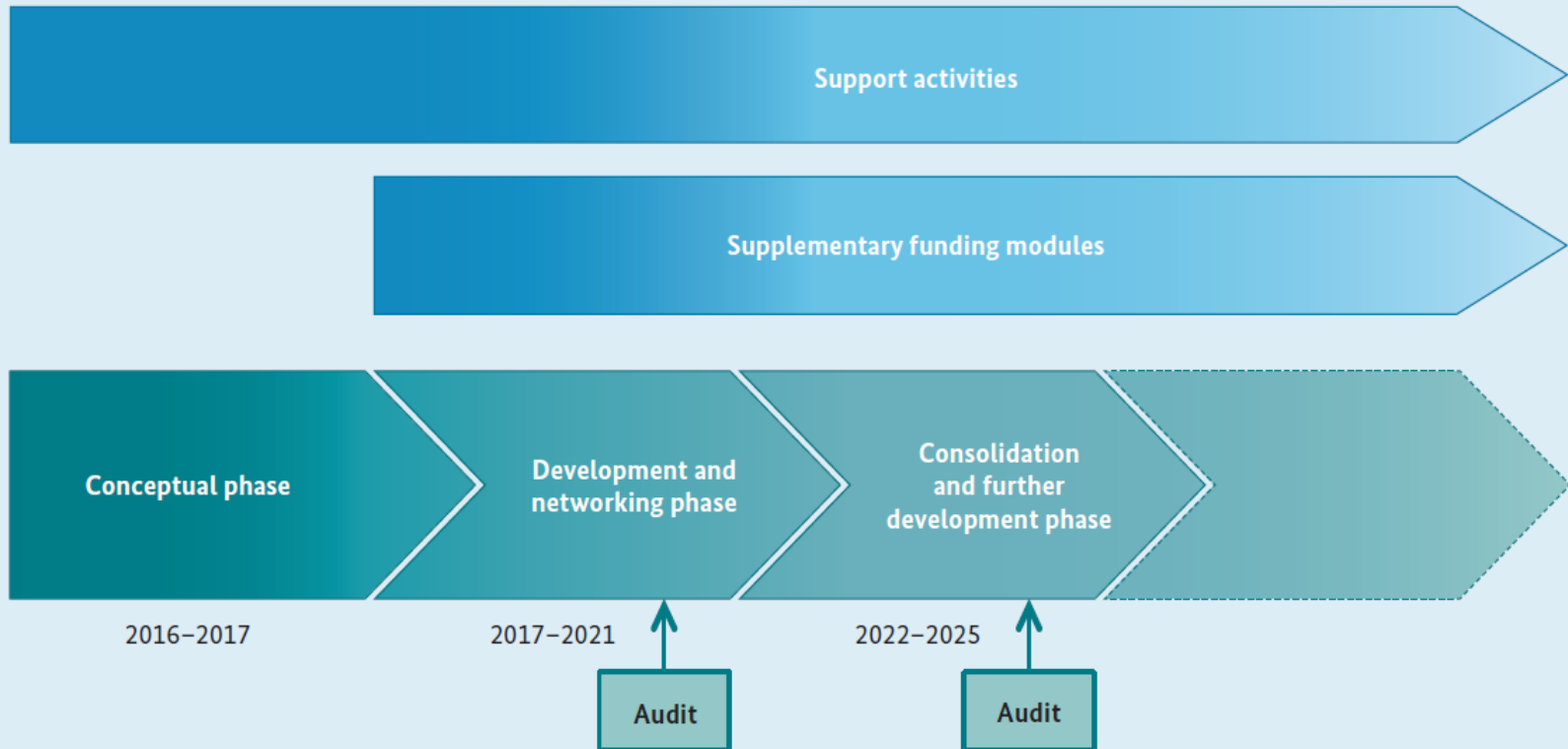


Federal Ministry  
of Education  
and Research

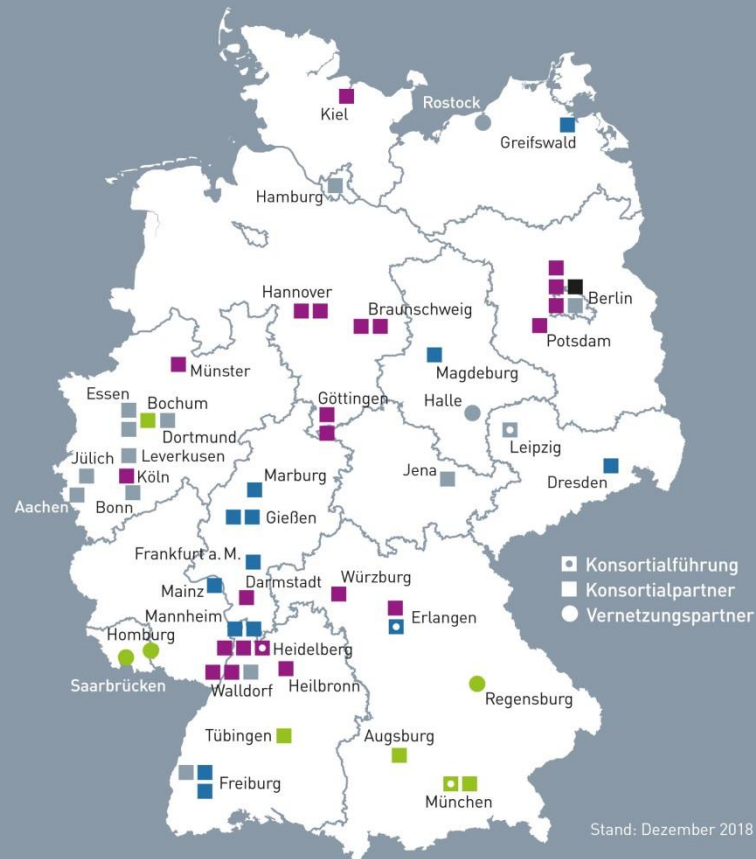
## Goals

- data exchange between, and joint data usage by, research and clinical care
- integration of data beyond the confines of individual institutions and locations
- improvement of both, medical research capabilities and clinical care
- strengthening of medical informatics in research, education and training

## Medical Informatics Funding Scheme



## Geförderte Konsortien und Standorte während der Aufbau- und Vernetzungsphase



DIFUTURE  
HiHmed  
MIRACUM  
SMITH  
Koordinationsstelle





# Summary

- Use of big data in medical research may be a **sensible supplement of**, but **not a substitute for**, hypothesis-driven and carefully designed scientific studies.
- Medical research depends upon the availability and usability of structured, high quality **patient-related data** (be they big or small)!
- Medical data integration in Germany is currently boosted notably by a large **governmental funding scheme** (the 'Medical Informatics Initiative').



*"Are we there yet?"*

*Thank you for your attention!*