

HelmholtzZentrum münchen

Deutsches Forschungszentrum für Gesundheit und Umwelt

OPEN SCIENCE – MEHR ALS NUR EINE FRAGE DER KOMMUNIKATION



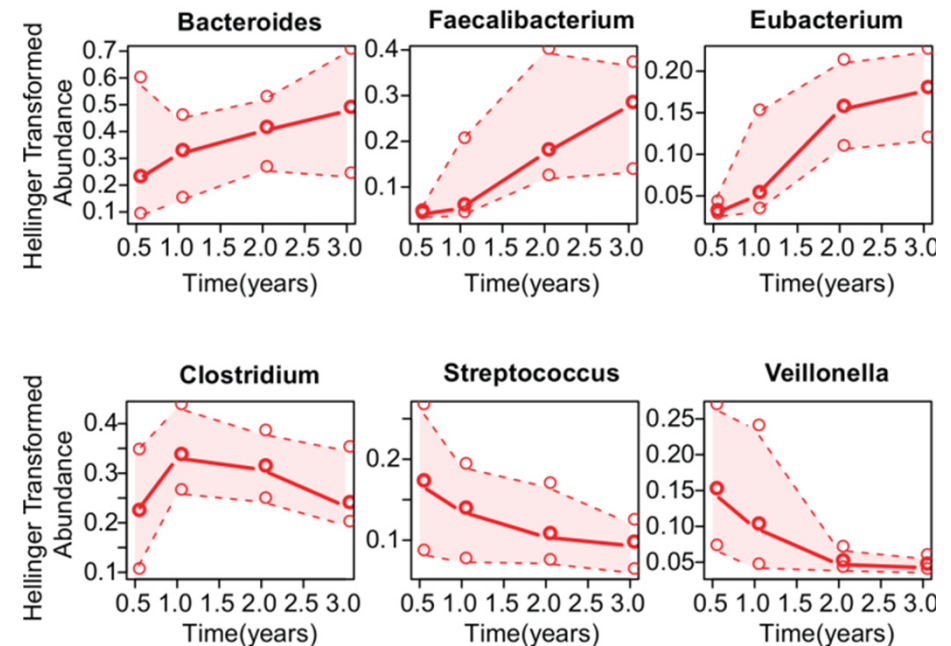
PD Dr. Wolfgang zu Castell

Helmholtz Zentrum München
Abteilung Scientific Computing

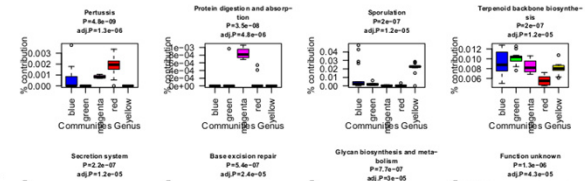
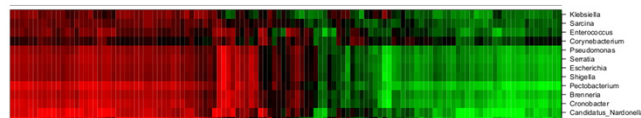
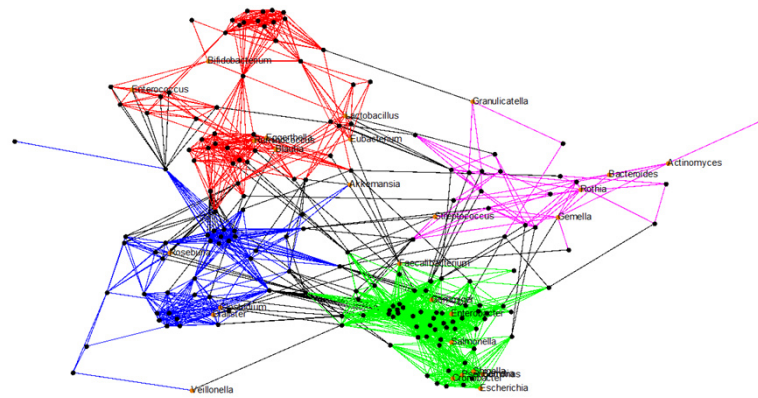
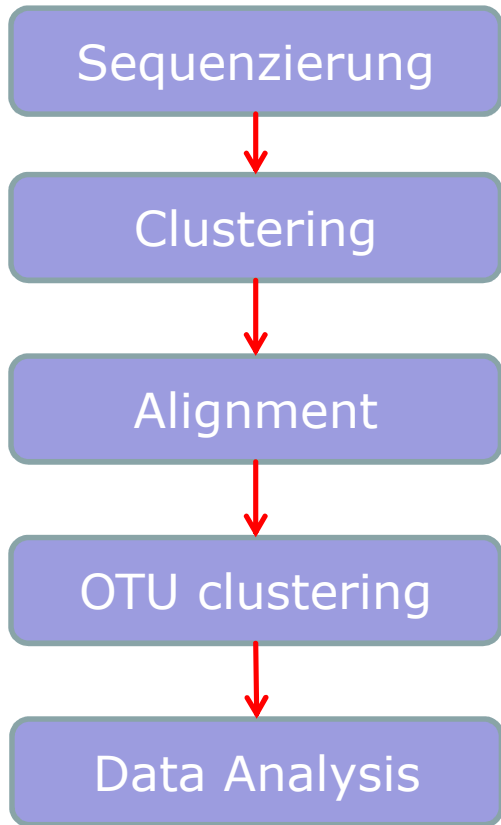
Köln, 08./09. September 2014

Ein Beispiel eines aktuellen Forschungsgebietes

- der Körper besitzt um einen Faktor 10 mehr Mikroorganismen als Körperzellen
- mehrere metabolische Störungen werden mit einer Veränderung der bakteriellen Gemeinschaft in Verbindung gebracht
- Mensch: 20.000-25.000 Gene
Mikrobiom: 3,3 Millionen Gene
- **die große Masse unserer Begleiter im Darm sind uns bisher unbekannt!**



Aktuelle Forschung nutzt massiv Informationstechnologie



2006 Diabetes Volume 63, June 2014

David Endesfelder,^{1,2} Wolfgang zu Castell,¹ Alexandria Ardisson,³ Austin G. Davis-Richardson,³ Peter Achenbach,² Michael Hagen,¹ Maren Pflueger,² Kelsey A. Gano,³ Jennie R. Fagen,³ Jennifer C. Drew,³ Christopher T. Brown,³ Bryan Kolaczowski,³ Mark Atkinson,⁴ Desmond Schatz,⁴ Ezio Bonifacio,^{5,6} Eric W. Triplett,³ and Anette-G. Ziegler²

Compromised Gut Microbiota Networks in Children With Anti-Islet Cell Autoimmunity

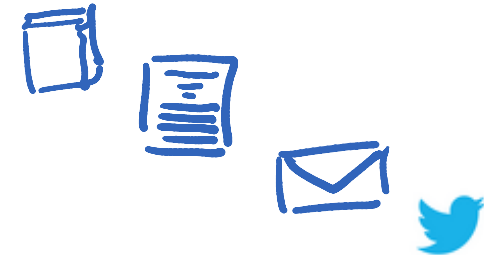
Diabetes 2014;63:2006–2014 | DOI: 10.2337/db13-1676

CrossMark

Informationstechnologie verändert die Gesellschaft ...

Peter Hinssen: „The **New Normal** -- when digital becomes normality“

Ein Buch kann über **Jahre** Relevanz besitzen;
ein Text-Dokument besitzt eine Relevanz von **Monaten**,
eine Email von **Tagen**;
ein Tweet veraltet nach **Minuten!**



Jede Organisation hat eine **innere Uhr**:

Wie schnell kann die Organisation auf Veränderungen ihrer Umwelt reagieren?

und eine **äußere Uhr**:

Wie schnell entwickeln sich die Kunden, kommunizieren und erwarten Antworten?

Digitalisierung bewirkt meist (zumindest gefühlt), dass ...

... die innere Uhr **langsamer** läuft, als die äußere Uhr.

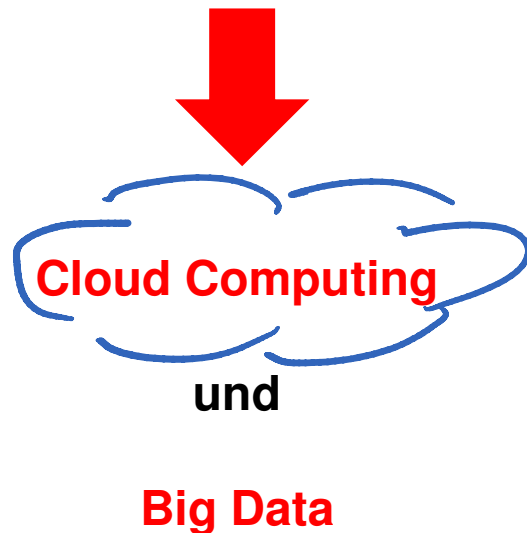
(nach P. Hinssen: The Era of Now, CA Technologies)

Die ‚Beschleunigung‘ erfordert eine Neuausrichtung

Hinszen formuliert zwei Grundregeln für die Informationstechnologie der „**Era of Now**“:

1. „Zero tolerance for digital failure“
2. „Good enough beats perfection“

(nach P. Hinszen: The Era of Now, CA Technologies)



Cloud Computing: Software- und Plattform-Applikationen aus ‚dem Internet‘ (SaaS, PaaS ...).

Big Data: zu groß um auf einem Rechner verarbeitet zu werden, weil die Datenmenge und/oder die Komplexität zu groß sind/ist.

Neue Möglichkeiten für die Forschung ...

Announcing the PLOS Text Mining

Collection

By Jyoti Madhusoodanan
Posted: April 17, 2013

However, as with the promises of many new technologies, the benefits of text mining are still not reaching most academic researchers.

This situation is now poised to change for several reasons. First, the **rate of growth of scientific literature has now outstripped the ability of individuals to keep pace with new publications** in a restricted field of study.

„the actual number of publications is less predictive of being cited than journal impact factors, suggesting that currently, the **perceived quality** of a publication ... is given more weight than its **actual quality**.“

Wissenschaft 2.0

Current Biology Vol 24 No 11
R516

Publication metrics and success on the academic job market

David van Dijk^{1,4}, Ohad Manor^{2,4},
and Lucas B. Carey^{3,*}

uses only the five most predictive features still has significant predictive power (AUC = 0.74; Supplemental information).

As expected, authors with more first author publications, and with more papers in high impact factor journals, are more likely to become PIs (Figure 1B). In addition, they have a higher *h*-index (*h* papers with at least *h* citations each), consistent with the

Aber ...

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding

Ioannidis, PLoS Medicine 2005

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False

Pro

Se

po

rat

co

is a

yet

...

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on **study power** and **bias**, the **number of other studies** on the same question, and, importantly, the **ratio of true to no relationships**.

Wo sind Probleme ... (zwei technische Beispiele)

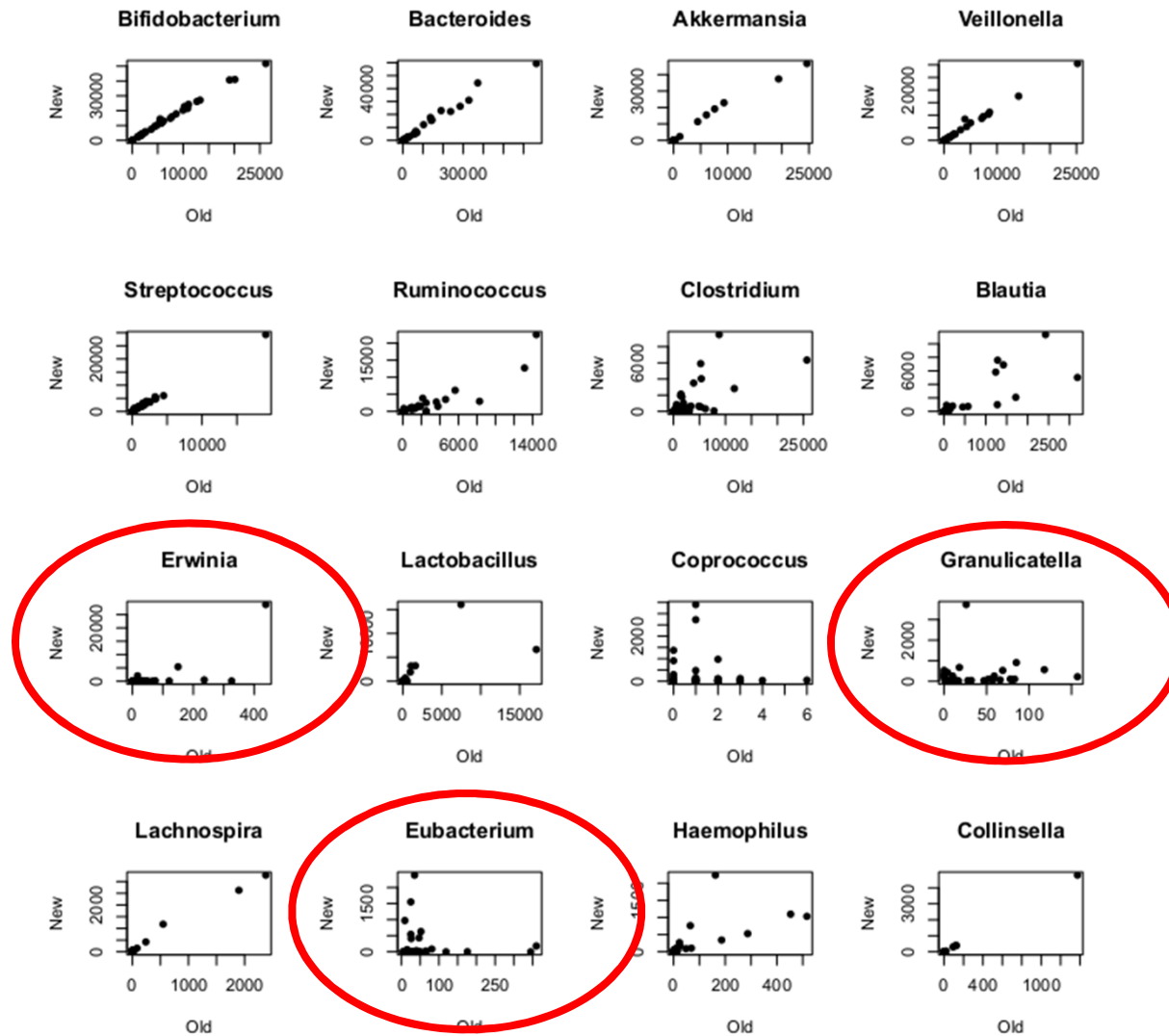


Abhängigkeit von Datenbanken

Beispiel:

taxonomische
Bestimmung
von OTUs

hier:
RDP vs. GG



Was messen wir wirklich?

Randy Schekman (Nobelpreis für Medizin 2013) in *The Guardian*, Monday 9 December 2013:

... **citation** is sometimes, but not always, linked to quality. A paper can become highly cited because it is good science – or because it is eye-catching, provocative or wrong. Luxury-journal editors know this, so they accept papers that will make waves because they explore sexy subjects or make challenging claims. **This influences the science that scientists do.** It builds bubbles in fashionably fields where researchers can make the bold claims these journals want, while discouraging other important work, such as replication studies.

Daniele Fanelli beispielsweise hat den „**Positive-Outcome-Bias**“ untersucht:

- Anteil der Berichte mit positivem Ergebnis ist von 70,2% (1990-91) auf 85,9% (2007) gestiegen
- Methodische Artefakte können diese und ähnliche Beobachtungen nicht erklären
- Es bleibt die Hypothese, dass *„research is becoming less pioneering, and/or that the objectivity with which results are produced and published is decreasing.“*

Fanelli, *Scientometrics* 2012

Risiken für die Wissenschaft

Der Unterschied zwischen Glauben und Wissenschaft liegt in der wissenschaftlichen Methode!

Anthony Chong 2013

„science“ that looks like science, but lacks the scientific integrity of utter honesty and repeatable experiments is little more than „**cargo cult science**“

The Dangers Of Cargo Cult Data Science, CIO Central, 4/01/2013

Richard Feynman 1974
(Nobelpreis für Physik 1965)



Cargo Cult Science

by RICHARD P. FEYNMAN

Some remarks on science, pseudoscience, and learning how to not fool yourself. Caltech's 1974 commencement address.

Engineering and Science, Volume 37:7, June 1974

Die ‚wissenschaftliche Methode‘ --- extrapoliert ...

- 1 Basis ist eine **Frage**
(zur Erklärung einer spezifischen Beobachtung oder auch eine offene Frage)
- 2 Ableitung einer **Hypothese**
(muss falsifizierbar sein)
- 3 **Prädiktion** auf Basis der Hypothese
(sollte die Hypothese von den wahrscheinlichsten Alternativen unterscheidbar machen)
- 4 **Testen** der Hypothese
(mögliche Fehler mind. besondere durch den Einfluß anderer wissenschaftlicher Kontrollen)
- 5 **Analyse** der gewonnenen Daten im Hinblick auf die Verifikation / Falsifikation der Hypothese

Open Science

Zukunftsperspektiven¹:

Science has always been driven by data but now the sheer complexity and amount of information is demanding new practices and methods to turn its bits into knowledge.

- **Open Peer-Reviewed** papers are **executable** as code
- **Wikipedia of science models** is created
- data-intensive science spawns new disciplines
- massively linked data becomes a **public utility**

¹ A Multiverse of Exploration. The Future of Science 2021, Institute for the Future, Fall 2011

Wie verhindern wir Cargo Cult Science?

Chong schlägt ein paar Maßnahmen vor:

- Verpflichtung zum sturen Festhalten an der Wahrheit
- Präsentation aller Ergebnisse, der positiven und der negativen
- insbesondere auch Laien dürfen nicht getäuscht werden
- ermögliche es jedem, eigene Schlüsse zu ziehen

The Dangers Of Cargo Cult Data Science, CIO Central, 4/01/2013

Verifizierbarkeit und **Reproduzierbarkeit** sind zwei der Grundpfeiler wissenschaftlichen Arbeitens, **Aufrichtigkeit** der wohl wichtigste.

Wissenschaftliche Qualität setzt voraus, dass über wissenschaftliches Arbeiten **offen**, **vollständig** und **unbedingt wahrheitsgemäß** berichtet wird.

Open Science ist damit nicht nur eine Frage der Kommunikation von Wissenschaft, sondern eine Frage der **Qualität zukünftiger Erkenntnis**.

*Es ist vielleicht heute noch nicht erkennbar,
wer morgen als Wissenschaftler bezeichnet wird,
aber es ist offensichtlich, woran wir ihn/sie
erkennen werden, sollten wir immer noch
von Wissenschaft sprechen.*

Vielen Dank für Ihre Aufmerksamkeit!