# Big Data

## Umgang mit personenbezogenen Daten im Kontext von digitalen Gesundheits-Akten und Archivierungslösungen

**Dr. med. Peter Langkafel MBA**

General Manager Public Sector / Healthcare MEE (Middle and Eastern Europe), SAP AG

Landesvorsitzender Berlin / Brandenburg (BVMI e.V.)

# Evolution von Autopiloten (Flugzeuge)

# Use Cases von Big Data in der Medizin

## Handlungsfelder von big data in der Medizin

- Regierung / Public Health
- Krankenkassen
- Care Provider (Krankenhäuser, Niedergelassene)
- Pharma
- Non traditional Player (Retail, Telco...)
- Versorgung (Einzelne Ärzte...)
- Forschung
- Patienten

- Health Education and Information
- Finanzielle Ressourcen-optimierung
- Public Health Monitoring
- Produkt-entwicklung)
- Prognose von Krankheiten
- Prävention
- Disease Management
- Optimierte klinische Studien)
- Konsumentenverhalten
- Medizinische Performance-Optimierung
- Informationiosaustausch (Patient zu Patient p2p)
- Kommunikation
- Adhärenz
- Neue Geschäftsfelder
- Entscheidungsunterstützung (clinical decision support)
- Schutz vor Missbrauch
- Kontrolle von Regulatorien
- Schnellere Implementierung von Evidenz
- Informationiosaustausch (Patient zu Doktor p2d)

Governance (Datenschutz, IT-Sicherheit, Informationelle Selbstbestimmung

© Langkafel

# Anatomy of Pavlok

ACTIVITY/SLEEP TRACKING

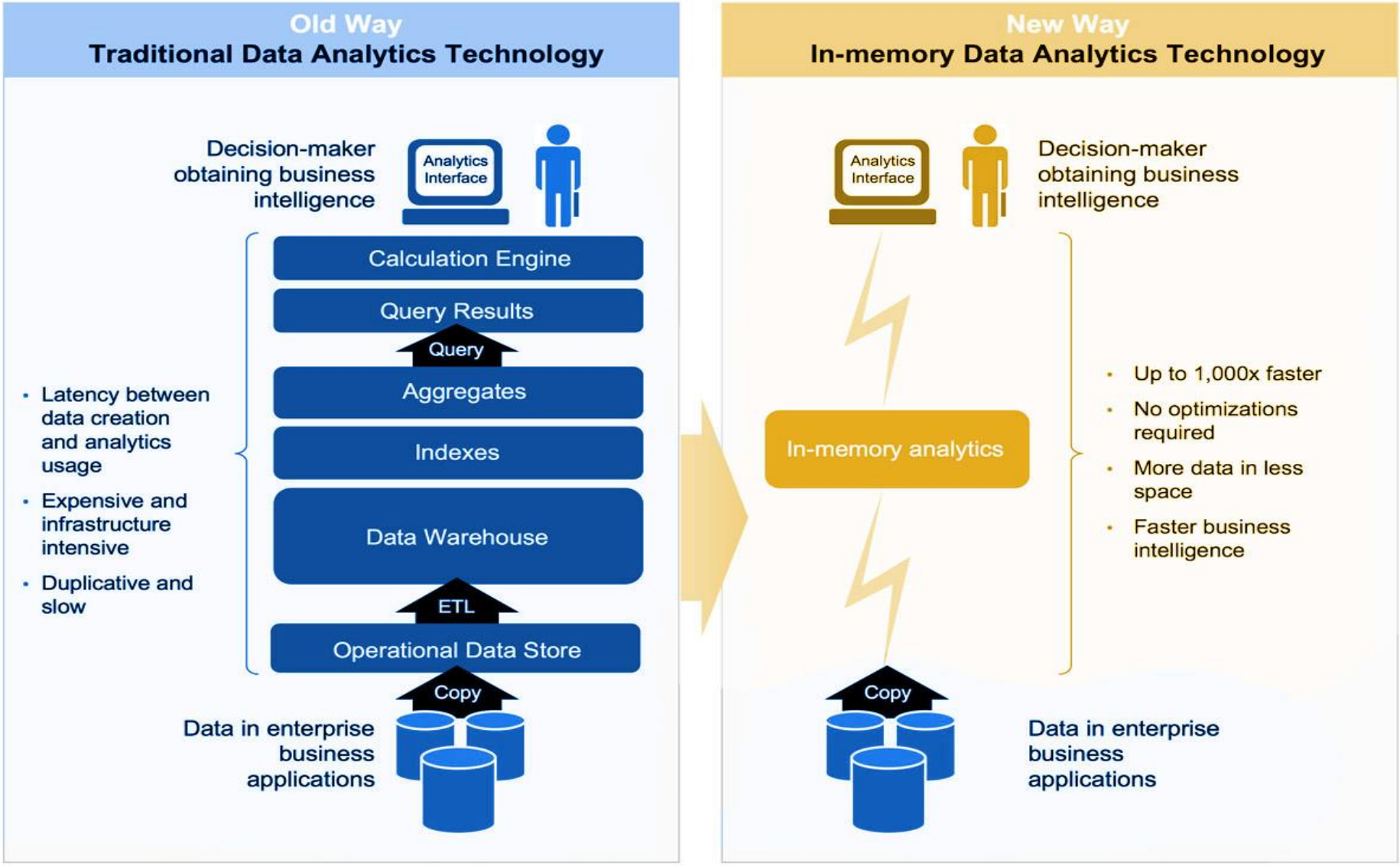
LOW ENERGY BLUETOOTH 4.0

REALTIME HAPTIC FEEDBACK

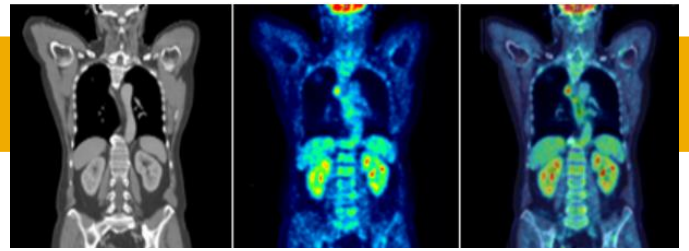
SHOCK CIRCUIT

HUMAN DETECTION

# SAP HANA – Simplification and Flexibility



# BIG DATA – good data – bad data?

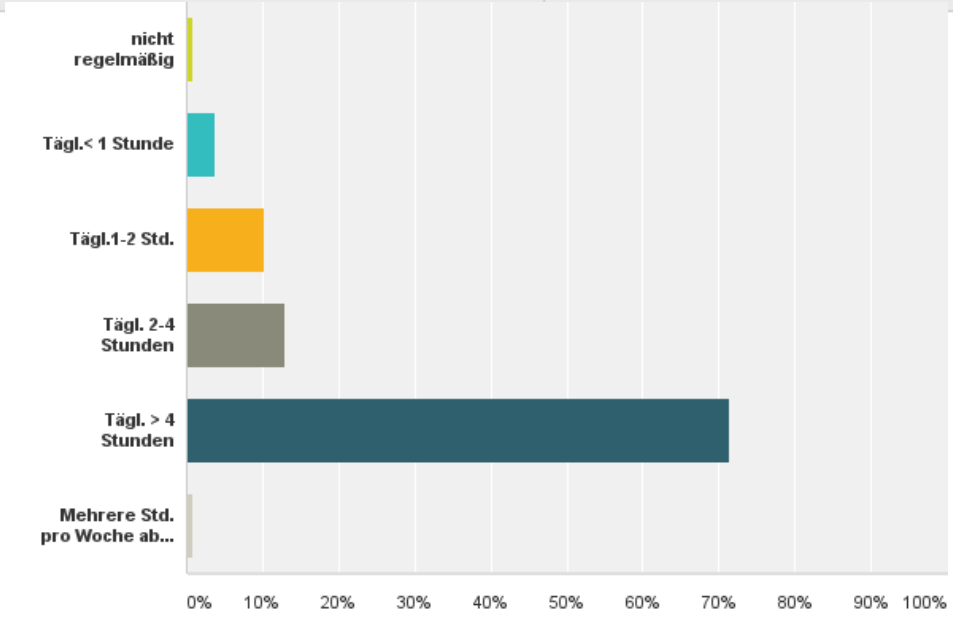


# Befragung Big Data in der Medizin: Februar – April 2015

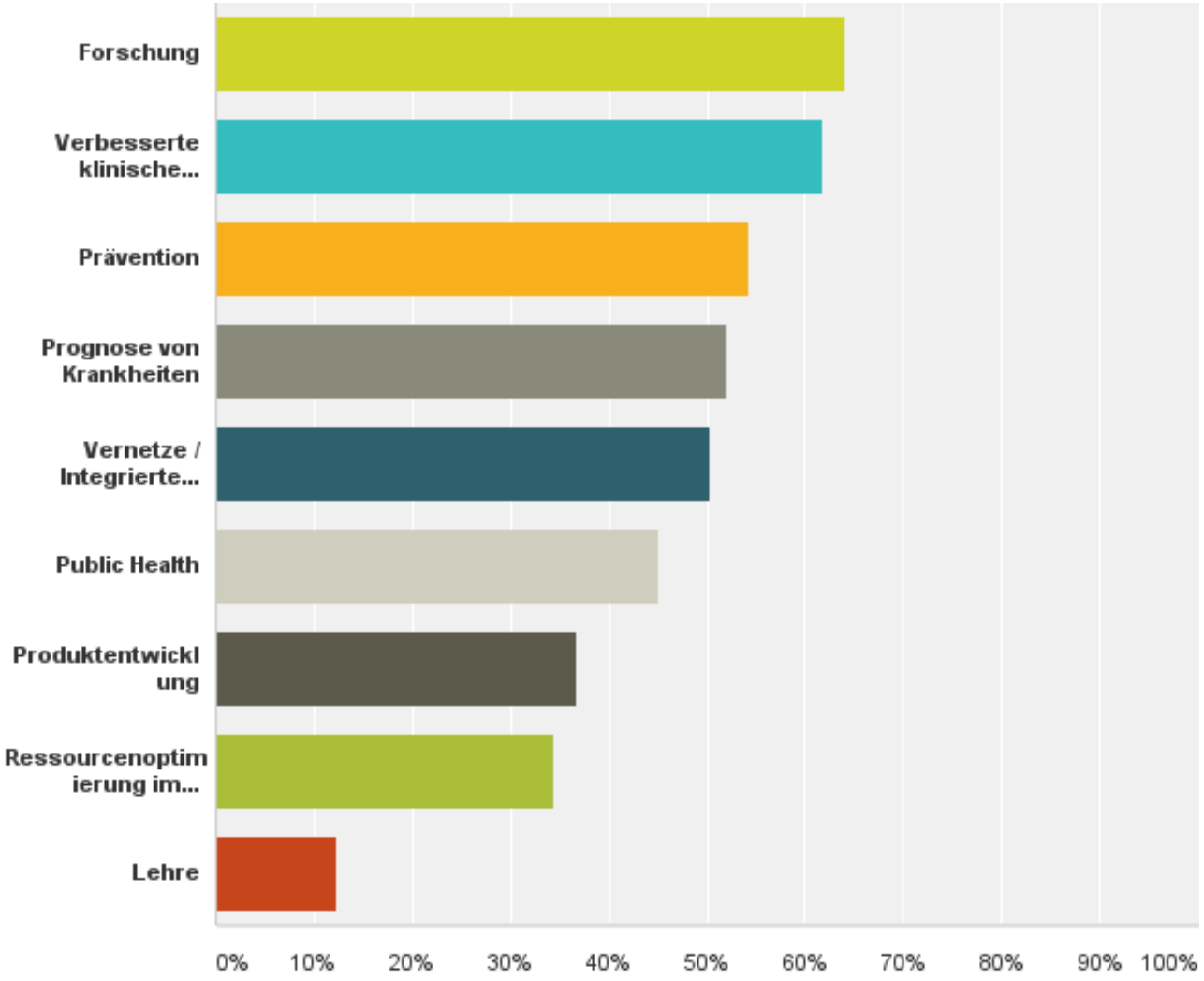
(Langkafel / Adam)

| Antwortmöglichkeiten | Beantwortungen | |
|---|---|---|
| Gesundheitspolitik / Verband | **21,37%** | 28 |
| IT Abteilung / Gesundheitswirtschaft | **14,50%** | 19 |
| Healthcare Beratung | **13,74%** | 18 |
| Management Gesundheitswirtschaft Krankenhaus | **10,69%** | 14 |
| IT - Industrie | **8,40%** | 11 |
| Ich bin beruflich nicht im Bereich Healthcare / Gesundheit tätig | **7,63%** | 10 |
| IT- Startup | **6,87%** | 9 |
| Pharma / Life Science | **6,87%** | 9 |
| Pflege | **3,82%** | 5 |
| Management Gesundheitswirtschaft Krankenkasse | **2,29%** | 3 |
| Medizintechnik | **2,29%** | 3 |
| Arzt in Klinik | **0,76%** | 1 |
| Arzt im Ambulanten Bereich | **0,76%** | 1 |
| **Gesamt** | | **131** |

| Antwortmöglichkeiten | Beantwortungen | |
|---|---|---|
| 18 bis 24 | **0,00%** | 0 |
| 25 bis 34 | **11,11%** | 12 |
| 35 bis 44 | **25,93%** | 28 |
| 45 bis 54 | **50,93%** | 55 |
| 55 bis 64 | **10,19%** | 11 |
| 65 bis 74 | **1,85%** | 2 |
| 75 oder älter | **0,00%** | 0 |
| **Gesamt** | | **108** |



# Wo sehen Sie großes Potential von Big Data in der Gesundheitswirtschaft?



# Big data – big Grauzone? WELCHE Daten?

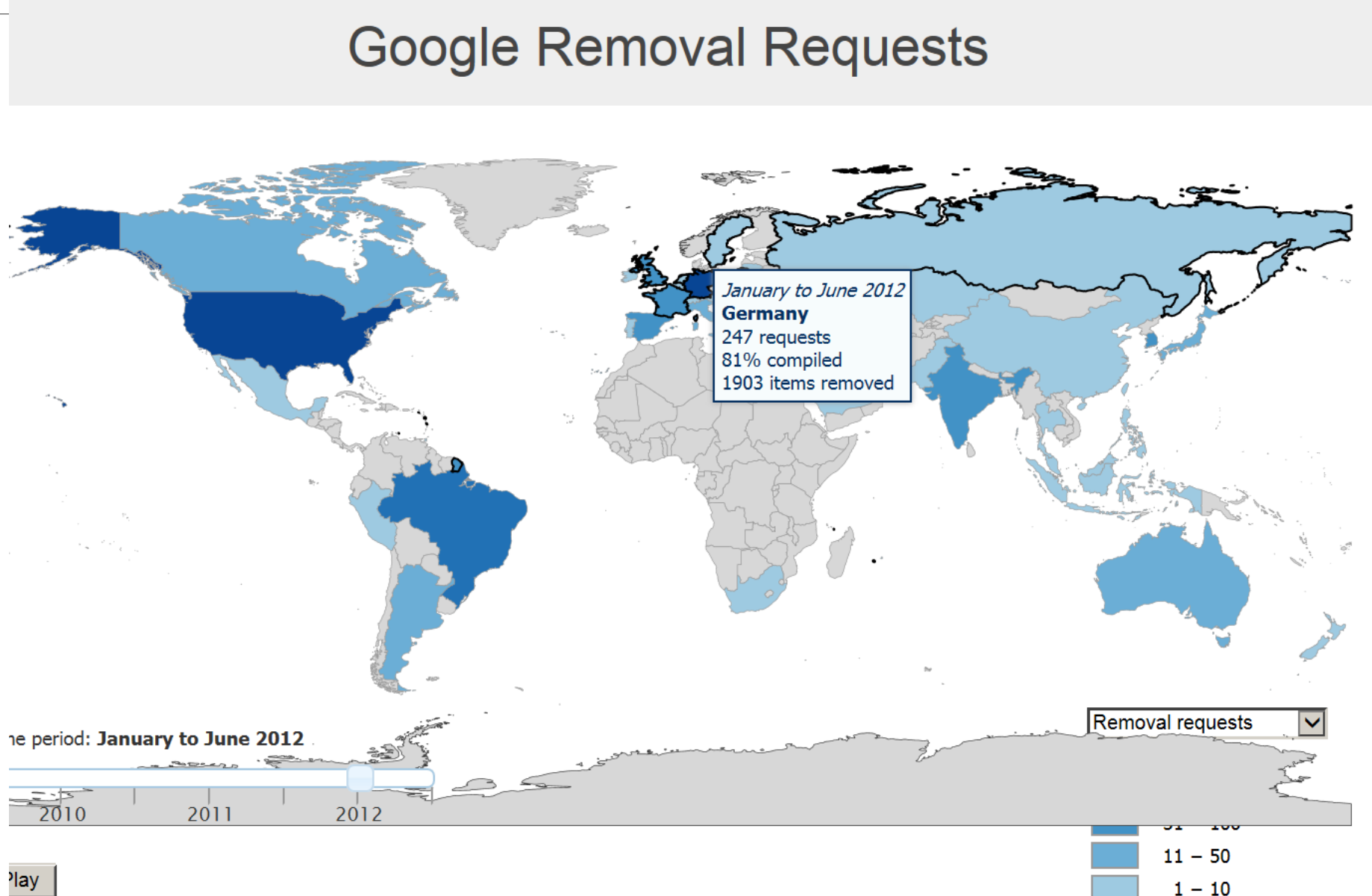
***Wir unterscheiden hier im Unternehmen zwischen „good“ und „bad“ Big Data. Und eigentlich ist diese Unterscheidung ganz einfach, sie richtet sich nämlich danach, worüber Daten gesammelt werden: geht es um ein Objekt, eine Sache – oder um eine Person?***

# T. Albig (Ministerpräsident S-H), Zeit März 2015

- Diskriminierungsfreie  Anzeige von Suchergebnissen
- Transparenzverpflichtungen für meinungsrelevante Algorithmen
- Wir sollten uns darüber klar werden, ob wir eine privilegierte Auffindbarkeit im Sinne eines *must be found* brauchen.
- Dem Nutzer sollten dem Presserecht entlehnte Rechte, wie zum Beispiel das auf Gegendarstellung, zustehen.
- Der Nutzer sollte ebenfalls erweiterte Informationsrechte gerade über die Nutzung seiner eigenen Daten erhalten.

# Big Data –Big Grauzone: Recht auf Löschung?

http://geography.oii.ox.ac.uk/?page=google-removal-requests



# Open Data – möglich / notwendig in der Medizin?!

Die 10 Prinzipien für offene Daten 33

1. Vollständigkeit

Sämtliche Rohdaten eines öffentlichen Datensatzes sollen in vollem Umfang zur Verfügung gestellt werden – bestenfalls unter Angabe von Formeln sowie Erklärungen zur Datenberechnung. Als öffentliche Daten werden hierbei Daten verstanden, die nicht berechtigten Datenschutz-, Sicherheits- oder Zugangsbeschränkungen unterliegen.

2. Primärquellen

Die Daten sollen an ihrem Ursprung gesammelt werden, und zwar mit dem höchstmöglichen Feinheitsgrad, also nicht in aggregierten oder anderweitig modifizierten Formaten.

3. Zeitliche Nähe

Daten werden so zügig wie zur Werterhaltung notwendig zur Verfügung gestellt – bestenfalls direkt nach deren Erhebung und Zusammenstellung.

4. Leichter Zugang

Daten werden so vielen Nutzern wie möglich für möglichst viele Verwendungszwecke bereitgestellt. Die gewünschten Informationen sollen hierbei so leicht als möglich eingeholt werden können.

5. Maschinenlesbarkeit

Um Daten automatisiert verarbeiten zu können, müssen die Informationen strukturiert in den weitverbreiteten Datenformaten zur Verfügung gestellt werden.

Open Knowledge Foundation

# Open Data – möglich / notwendig in der Medizin?!

6. Diskriminierungsfreiheit

Ohne vorherige Identifikation (Registrierung oder sonstige Mitgliedschafts-Voraussetzungen) sollen die Daten jedem zu jeder Zeit zur Verfügung stehen.

7. Verwendung offener Standards

Daten werden in standardisierten, kostenlos verfügbaren Formaten bereitgestellt, über die keine juristische Person die alleinige Kontrolle hat.

8. Lizenzierung

Daten unterliegen keinen Nutzungsbeschränkungen wie Urheberrecht, Patent, Markenzeichen oder Geschäftsgeheimnis. Sinnvolle Datenschutz-, Sicherheits- und Zugangsbeschränkungen sind jedoch zulässig.

9. Dauerhaftigkeit

Öffentliche Datenströme sollen dauerhaft in Online-Archiven verfügbar gemacht werden.

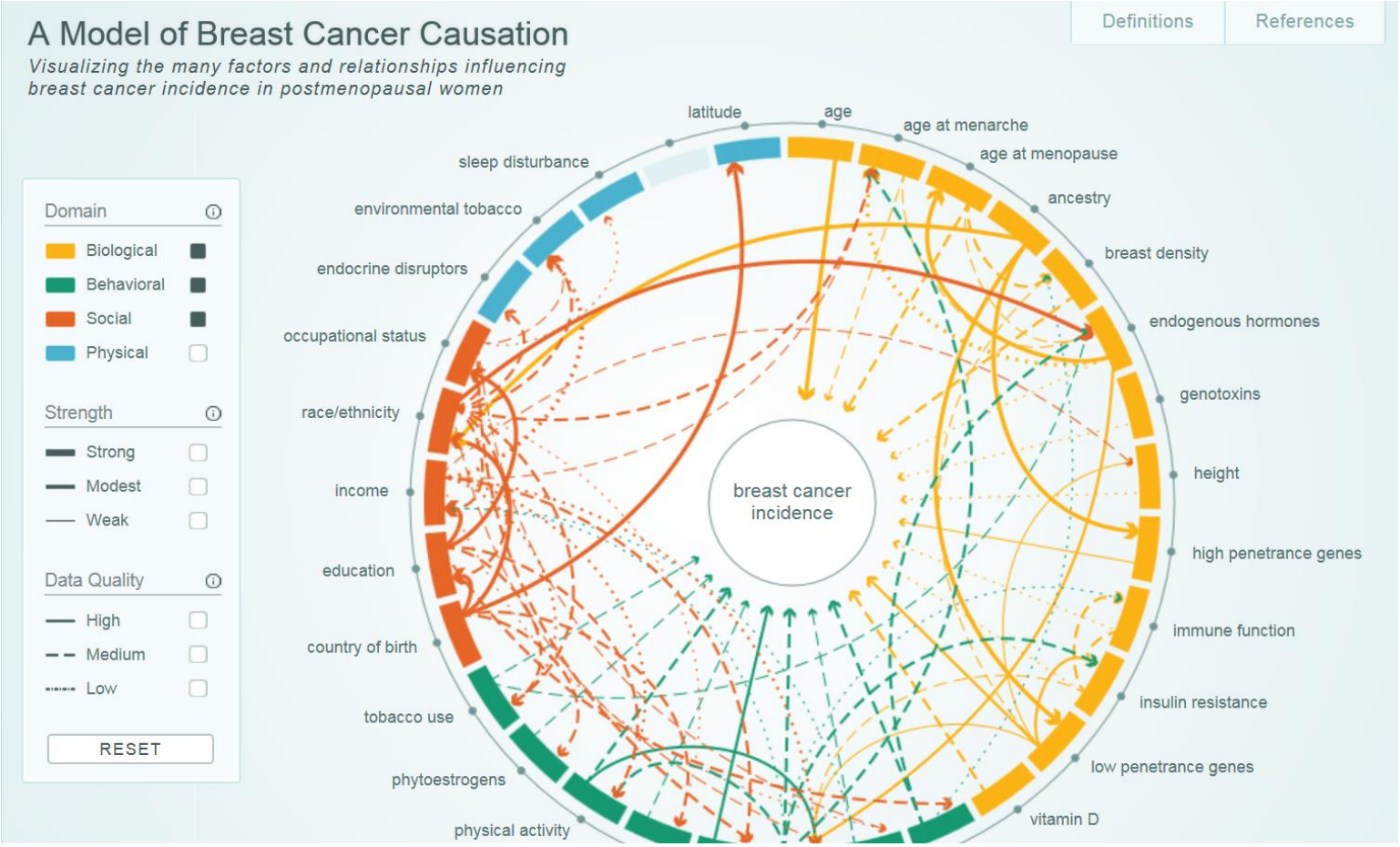
10. Niedrige Nutzungskosten

http://sunlightfoundation.com/blog/2011/07/14/vivek-kundras-10-principles-for-improving-federal-transparency/.

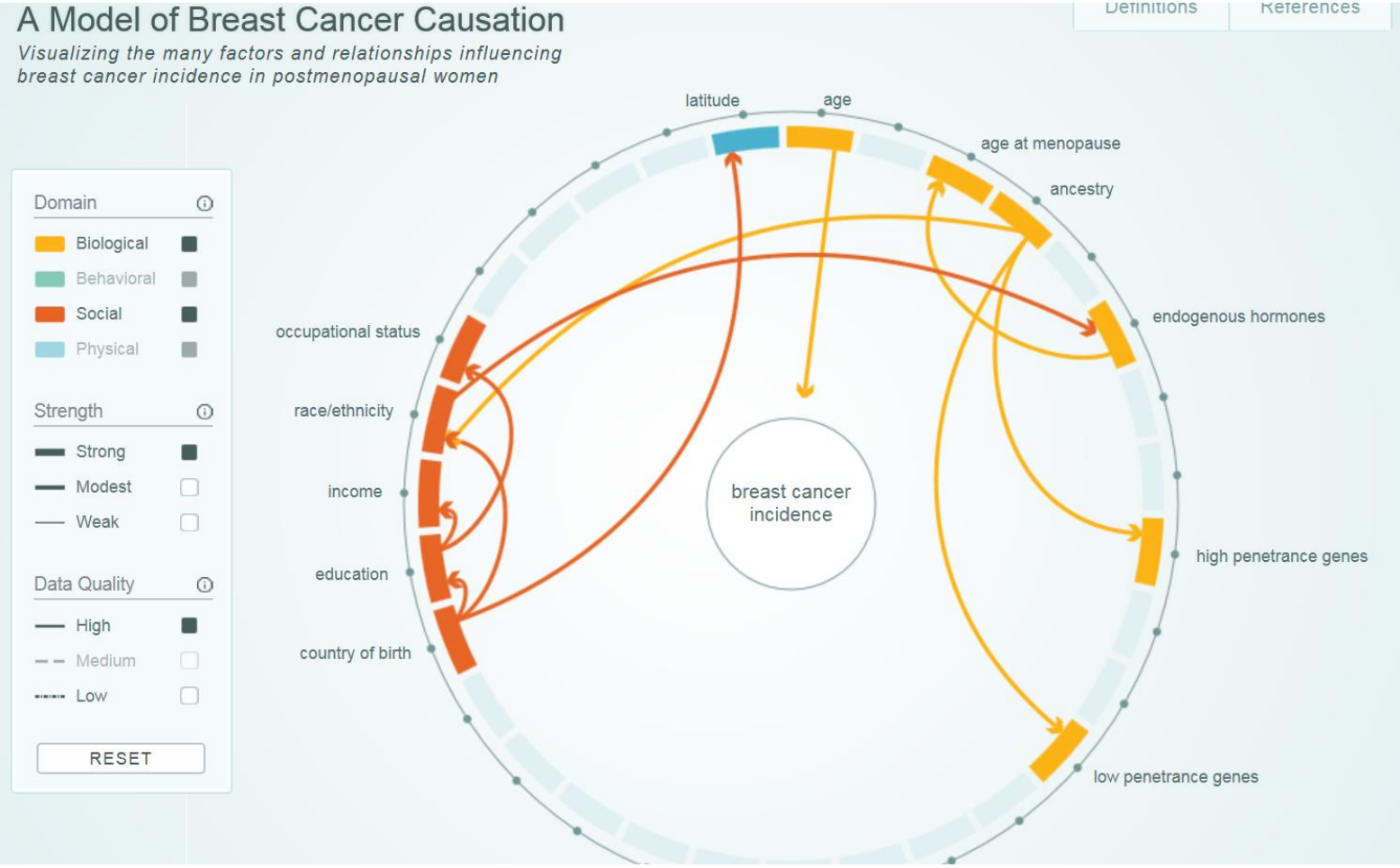
http://wiki.okfn.de/10-Prinzipien-fuer-offene-Daten

Open Knowledge Foundation

# Visualisierung
# http://www.cabreastcancer.org/causes/



# Beispiel


A Model of Breast Cancer Causation
Visualizing the many factors and relationships influencing breast cancer incidence in postmenopausal women
Definitions
References
Domain
Biological
Behavioral
Social
Physical
Strength
Strong
Modest
Weak
Data Quality
High
Medium
Low
RESET
latitude
age
age at menopause
ancestry
endogenous hormones
high penetrance genes
low penetrance genes
occupational status
race/ethnicity
income
education
country of birth
breast cancer incidence

# Big Data

## Umgang mit personenbezogenen Daten im Kontext von digitalen Gesundheits-Akten und Archivierungslösungen

**Dr. med. Peter Langkafel MBA**

General Manager Public Sector / Healthcare MEE (Middle and Eastern Europe), SAP AG

Landesvorsitzender Berlin / Brandenburg (BVMI e.V.)

# SAP HANA – Enabler in Memory Computing

| Innovation | Benefit | Application |
|---|---|---|
| **Multi-core architecture** | Massively parallel execution | High throughput sequencing analysis |
| **6 TB DRAM servers** | Large Data Sets in-memory | Genomics, proteomics and patient data |
| **Compression (5-20x)** | Large data sets in-memory | Genomics, proteomics and patient data |
| **Combined Column and Row Store** | Column = Fast Queries | Adhoc queries using clinical data |
| **Partitioning: In-Database computing** | Analyze large data sets<br>Complex computations | Genome alignment<br>Proteomics and Imaging data |
| **No aggregate tables** | Flexible modeling<br>No data duplication | Data Model for combined clinical and omics data |
| **Text Analytics** | Use of unstructured data | Physician's letters<br>Scientific Literature |