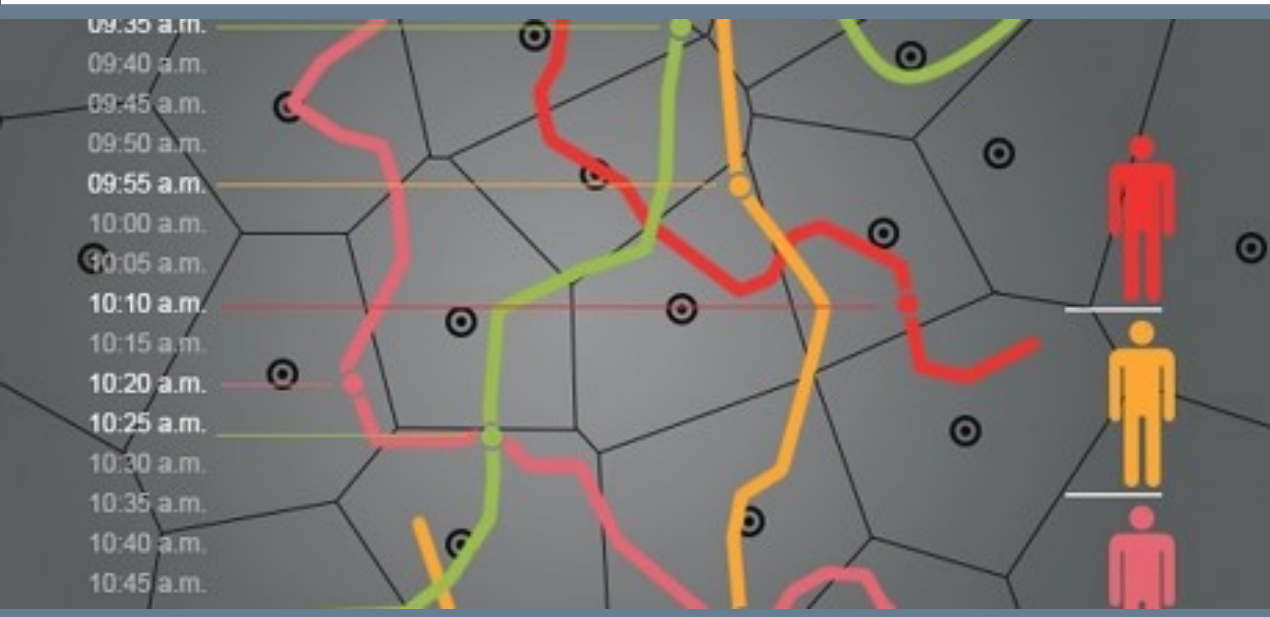




Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences



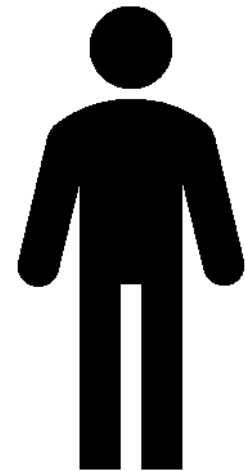
Anonymisierung und Big Data

Prof. Dr. Reinhard Riedl
BFH-Zentrum Digital Society

Anonymisierung



- Den Personenbezug von Daten eliminieren
 - Durch Löschen / unlesbar machen von Namen oder von anderen eindeutigen oder identifizierenden Merkmalen
 - Durch Verschlüsseln von Daten



De-Anonymisierung

- Den Personenbezug von Daten wieder herstellen
 - Selten: 99.9999..% Sicherheit
 - Häufig: 90 – 99%
 - Oft nur: genügend hoch für personalisierte Preise und Werbung



Rolle des Kontexts

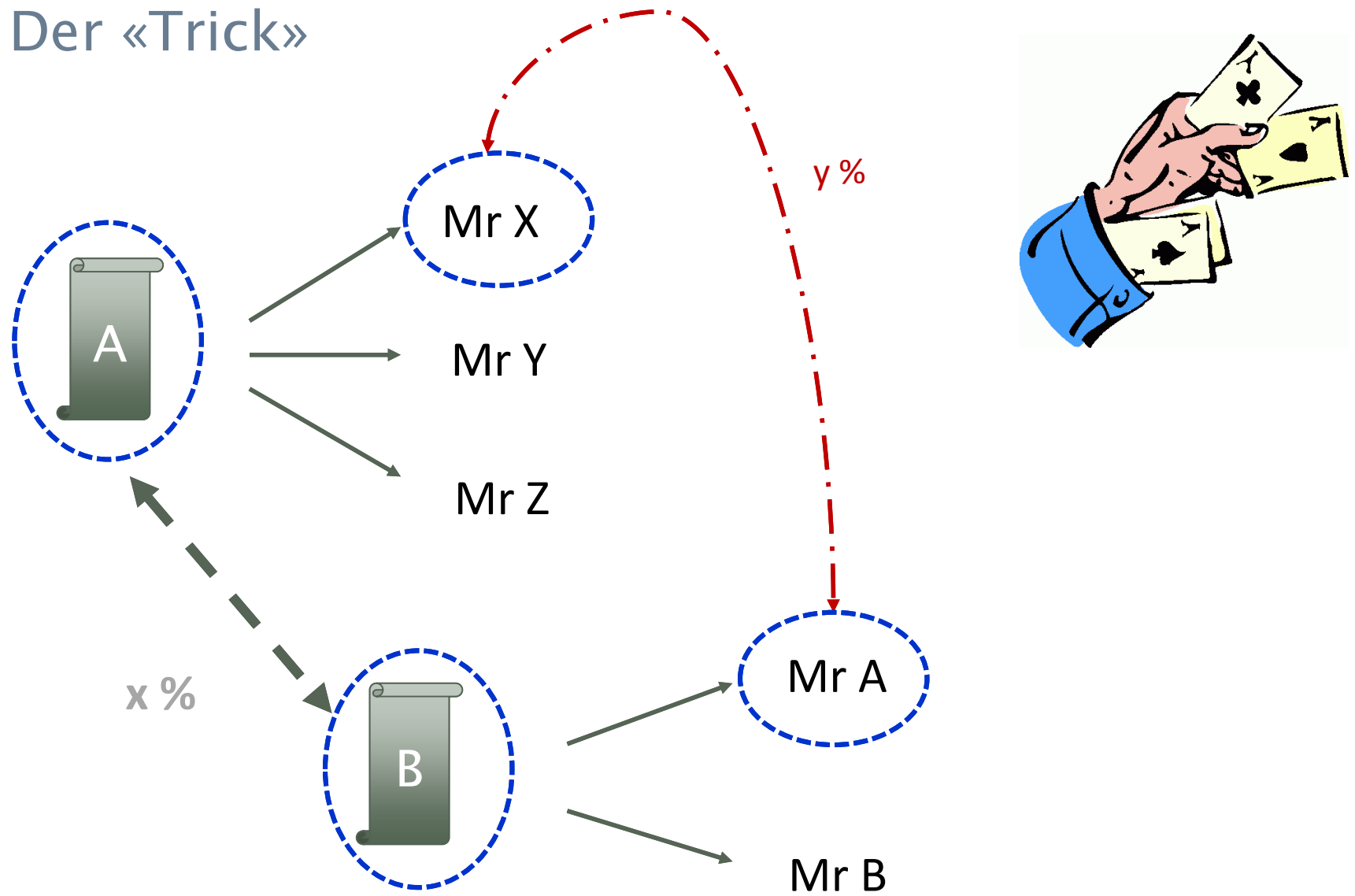


- Je enger definiert der Kontext ist
 - Desto mehr identifizierende Merkmale gibt es
 - Desto weniger Information ist zur Identifikation notwendig

- Je mehr Daten es zum Kontext gibt
 - Desto enger ist der Kontext
 - Desto mehr identifizierende Merkmale gibt es
 - Desto weniger Information ist zur Identifikation notwendig

- Nicht personenbezogene Kontext-Daten können zur De-Anonymisierung führen

Der «Trick»



Die Kernelemente der De-Anonymisierung

1. Datenobjekte in reichhaltigem Kontext betrachten
2. Kontextbezogen Datenobjekte mit Eigenschaften «korrelieren»
3. Datenobjekte eigenschaftsbasiert untereinander «korrelieren»
4. Über Personen Daten sichere und wahrscheinliche sammeln
5. Datenobjekte mit Personen «korrelieren»

Ergebnis: Pseudo-De-Anonymisierung:

nutzbar für kommerzielle und kriminelle Aktivitäten, politische und andere Kampagnen, etc.

BIG DATA =

**implizit in Daten vorhandene Informationen
explizit machen**

**aus Daten etwas herauslesen, wofür sie nicht
gesammelt wurden**

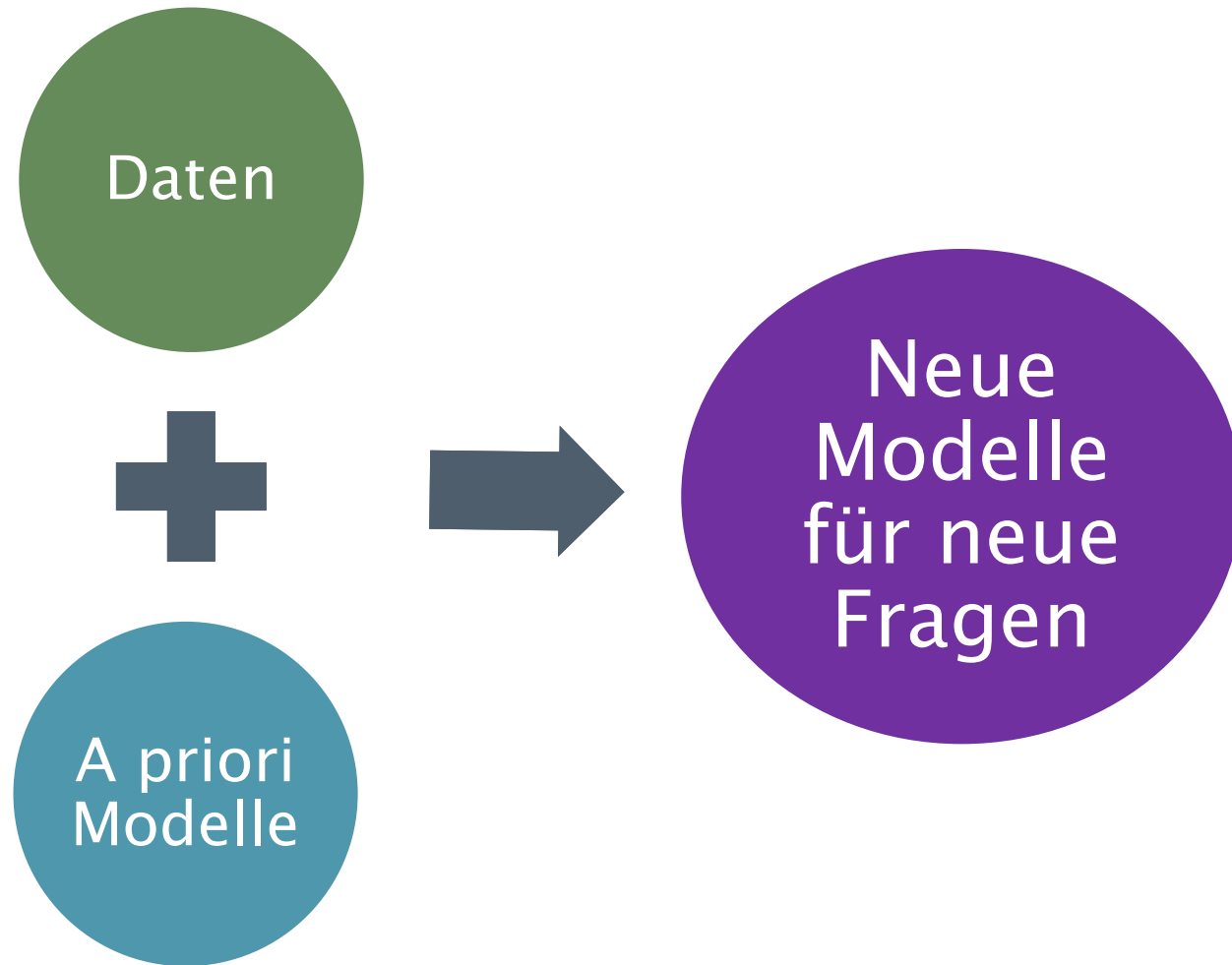
Typischer Fall

- INPUT:
Viele Daten, bunt gemischt und von zweifelhafter Qualität, teils weiss man «in etwa», was man hat, teils muss man raten ...
- ANGESTREBTER OUTPUT:
Resultate, deren Evidenz-Qualität gut eingeschätzt werden kann, je nach Situation und Verfahren aber variiert
 - Von Hypothesen
 - ...
 - bis zu kausalen Zusammenhängen


$$0 + 0 + 0 + \dots + 0 = 1$$

Viele anekdotische Einzelfällen → Evidenz

Der Megatrück: Daten liefern Auswertungsmodelle!



Wichtiges Einsatzszenario für De-Anonymisierung: Schätzen von aus anonymen Daten ableitbaren Eigenschaften



Daten **von vielen**
darüber, wie sie sich
in den Kontexten
A und **X** verhalten

?

←
Schätzung
bzw. Prognose
→




Das Verhalten **einer**
Person im **Kontext A**
ist bekannt, wir
interessieren uns für
ihr Verhalten im
Kontext X

*Big Data hilft, unbekannte Eigenschaften zu schätzen,
bzw. Prognosen für Verhalten in Kontexten zu machen*

Weiteres Einsatzszenario für De-Anonymisierung: Finden von Mustern



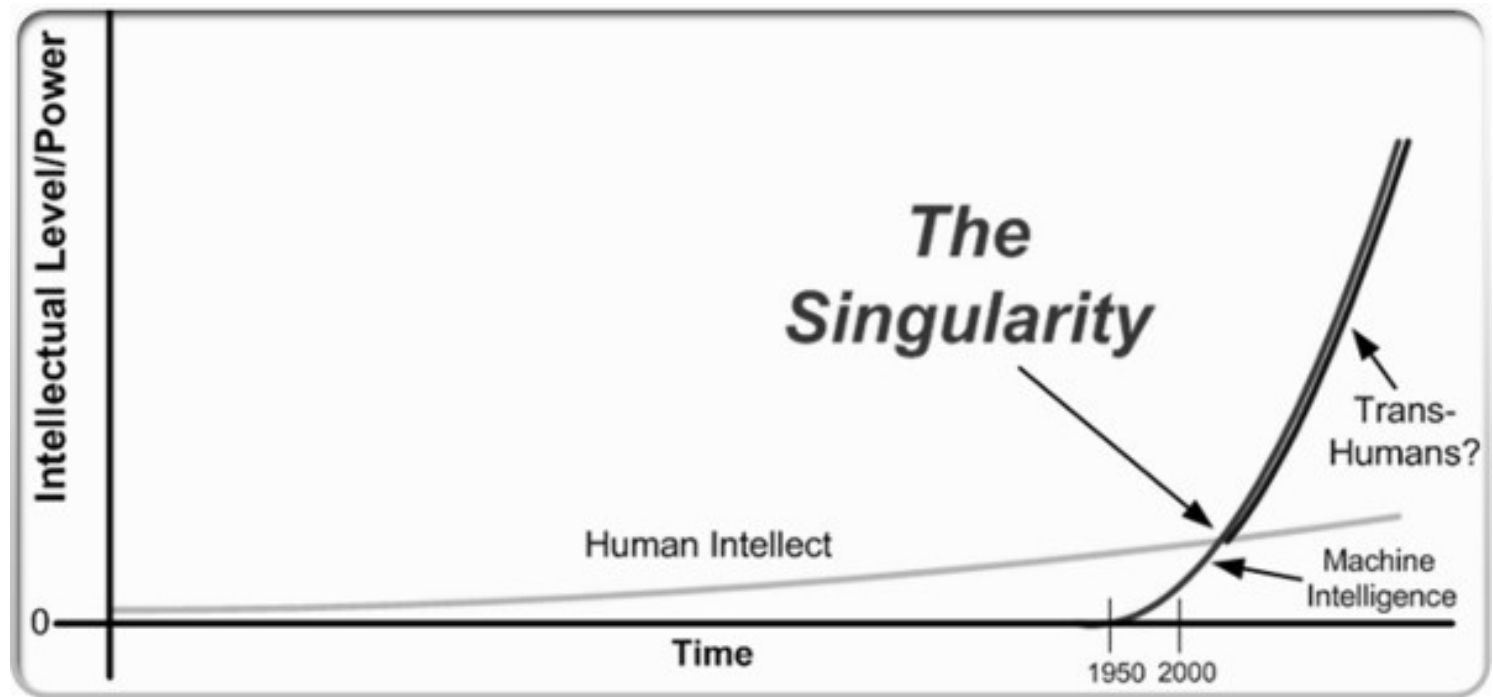
Daten **von vielen**
in unterschiedlichen
Kontexten

Identifikation eines
untypischen, aber

für eine Person
charakteristischen
Musters



**Big Data macht aus dem «Trick»
eine Maschinerie!**

Die bedrohliche Hoffnung der «Dataisten»



Saravanan.org

«*Datenwissenschaft ersetzt andere Wissenschaften!*»

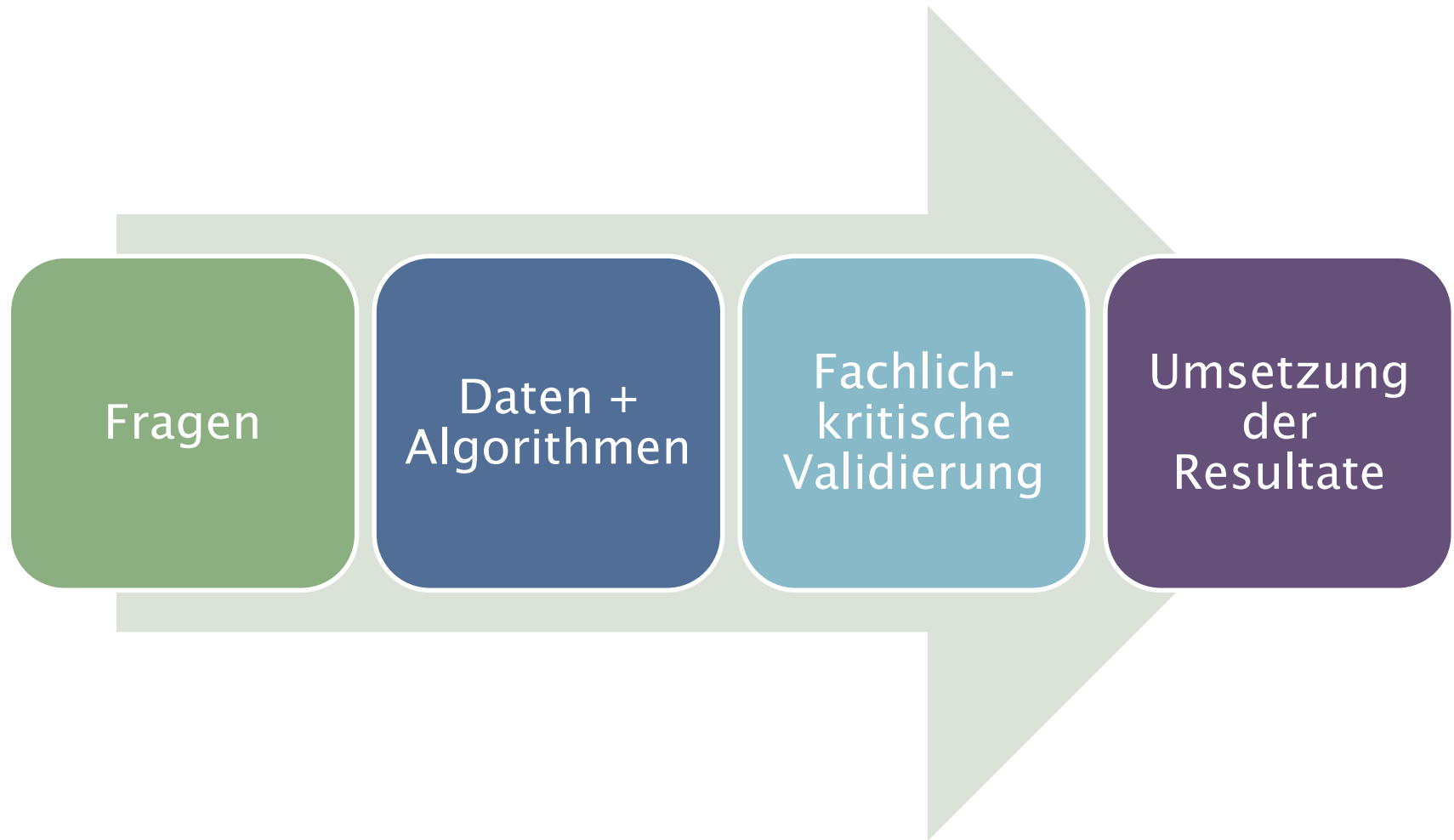
das galt 2008 – 2015 ... mittlerweile ist klar:

... in Einzelfällen funktioniert das schon!

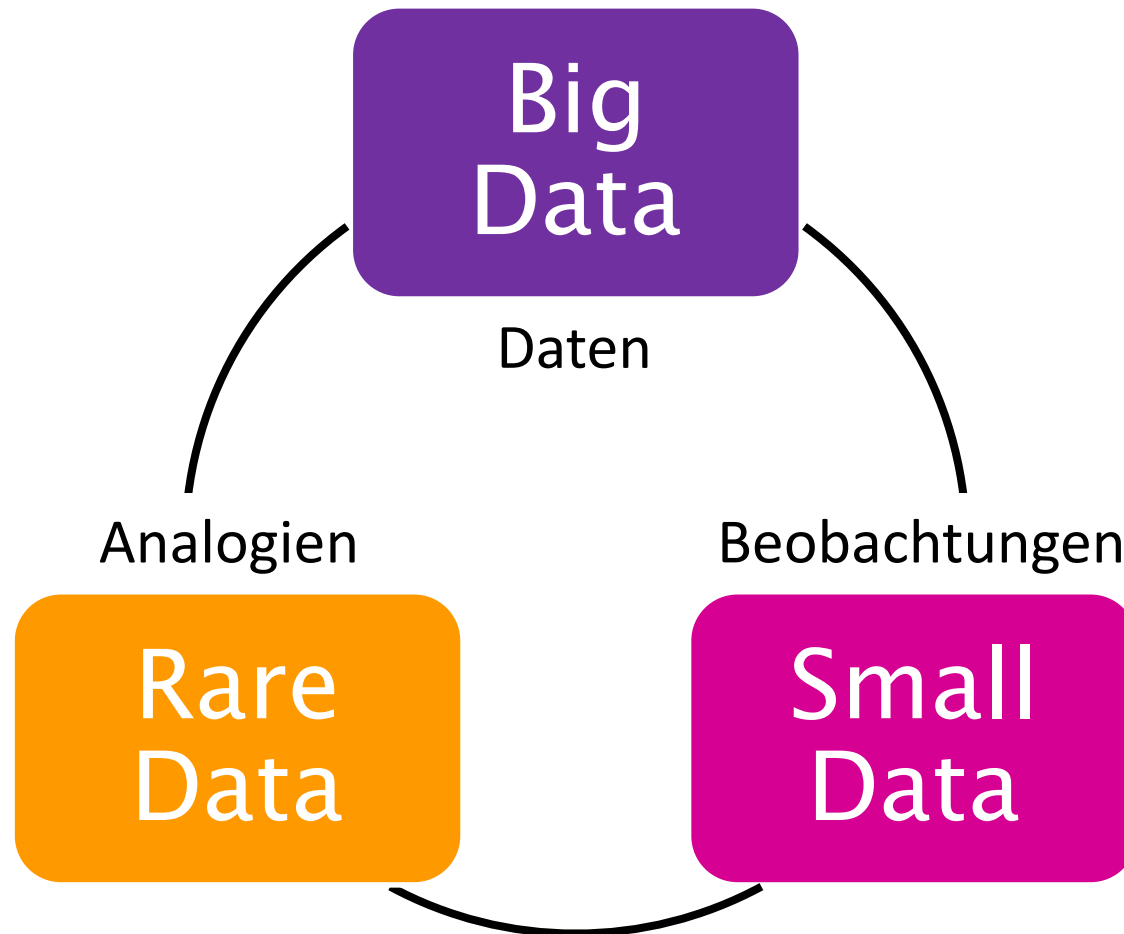
BIG DATA unterstützt Pseudo-De-Anonymisierung, ... und verführt zur **Pseudo-Pseudo-** De-Anonymisierung!

1. Mit diffusen **Fragen**,
gefährlich missverständliche Antworten!
2. Ohne **kritische Validierung** der Antworten,
keine verantwortungsvoll nutzbaren Antworten!
3. Ohne Bereitschaft **zur stochastischen Rechnung**
grober Unfug!

Die Big Data Pipeline muss beherrscht werden



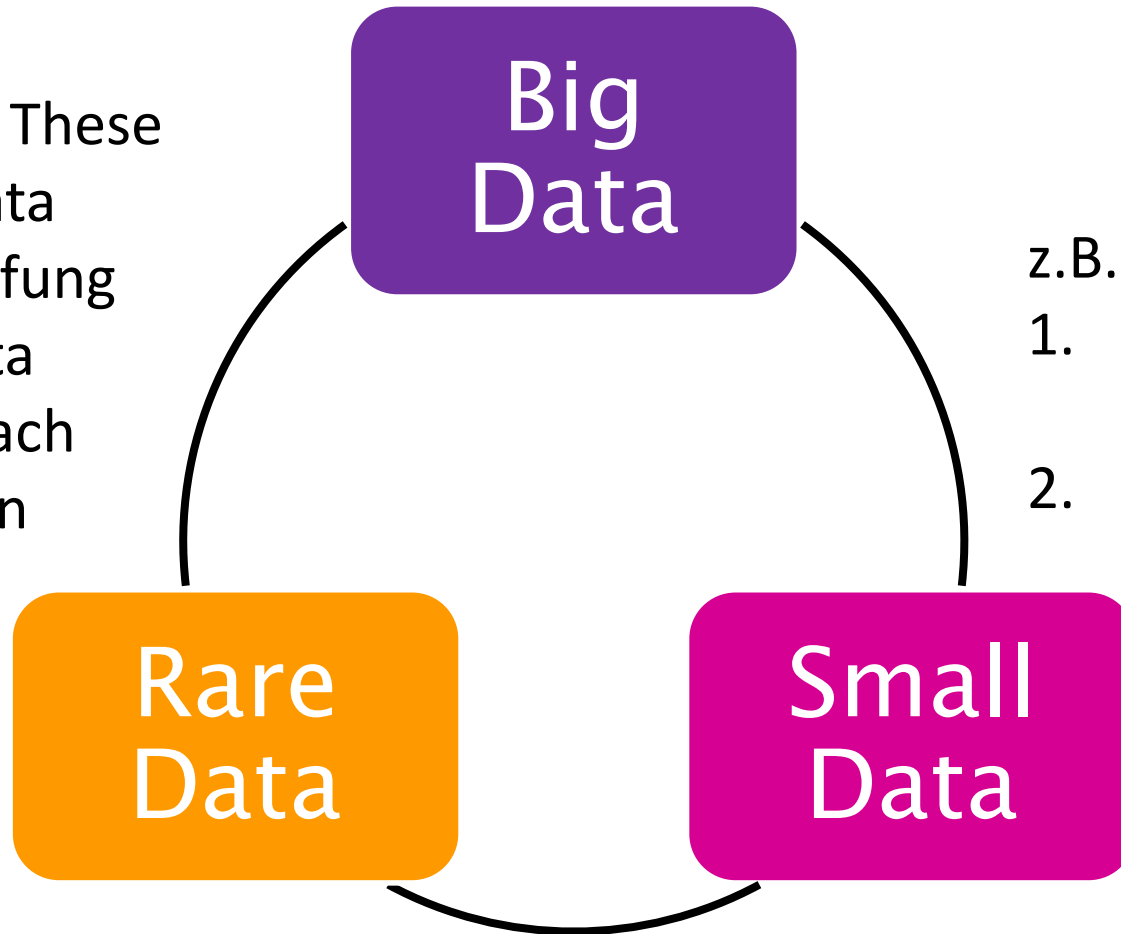
Big Data ist zudem nicht die einzige Erkenntnismethode



In der praktischen Arbeit sind alle drei involviert ...

z.B.

1. Big Data These
2. Small Data Überprüfung
3. Rare Data Suche nach Lösungen



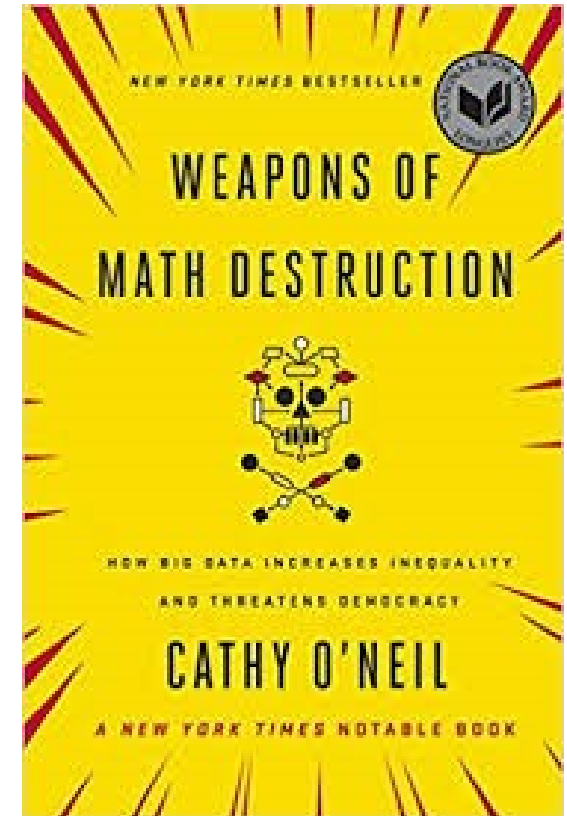
z.B.

1. Small Data Thesen
2. Big Data Tests

- z.B. 1. Rare Data Idee für neue Dienste
2. Big Data Detailanalysen

In den Händen von Unwissenden und Böswilligen

- Gefährlich sind Modelle mit 4 Eigenschaften
 - Fehlende Transparenz
 - Kein Feedback-Mechanismus
 - Hohe Skalierbarkeit
 - Grosse negative Auswirkungen auf Betroffene
- Der Schutz der Privatsphäre ist eine Illusion
 - Z.B. dauerhafte Anonymisierung
- Viele Anwendungen sind per se gesellschaftlich schädlich





Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

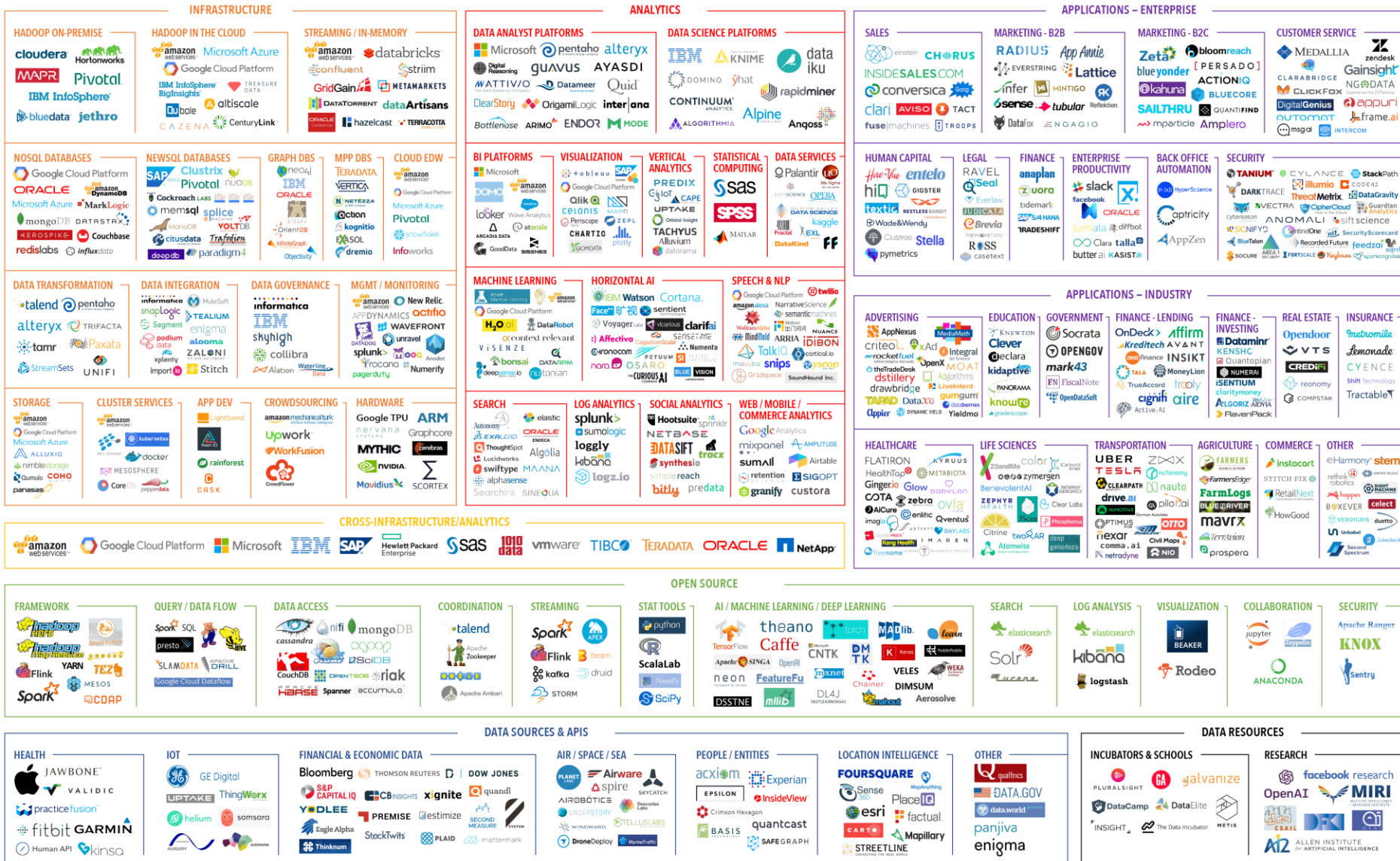
Fazit

Anonymisierung ist Glatteis

- Verschlüsselungen sind nicht dauerhaft
 - Blockchain ist ein Synonym für «Alles kommt auf!»
- Anonymität ist eine instabile Eigenschaft
 - Was heute gilt, muss nicht in Zukunft gelten
 - Das Ende ist nur für den Anonymitätsbrecher erkennbar
 - Je mehr Daten, desto unsicherer die zukünftige Anonymität
- Daten und Algorithmen müssen verstanden werden!
- Kritisches Denken ist notwendig!
- Die Mathematik hilft ein wenig!



BIG DATA LANDSCAPE 2017





Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

Danke! Fragen?

reinhard.riedl@bfh.ch

BFH-Zentrum Digital Society



- 6 Fakultäten (Departemente)
 - Wirtschaft; Technik & Informatik; Hochschule der Künste Bern; Gesundheit; Soziale Arbeit; Architektur, Holz und Bau;
- 6 Themen
 - 2 Methodenthemen: Nachhaltiges Design; Angewandte Datennutzung;
 - 2 Fachthemen: Identität & Privatsphäre; Cybersecurity & IT-Forensik;
 - 2 Anwendungsthemen: Gebäude & Städte; Gesundheitsversorgung;
- Mehr als 12 Disziplinen
- Eine Online-Zeitschrift
 - www.societybyte.swiss
- Eine Konferenz
 - Transform – Digital Transformation Skills
 - 13. September, Berner Rathaus