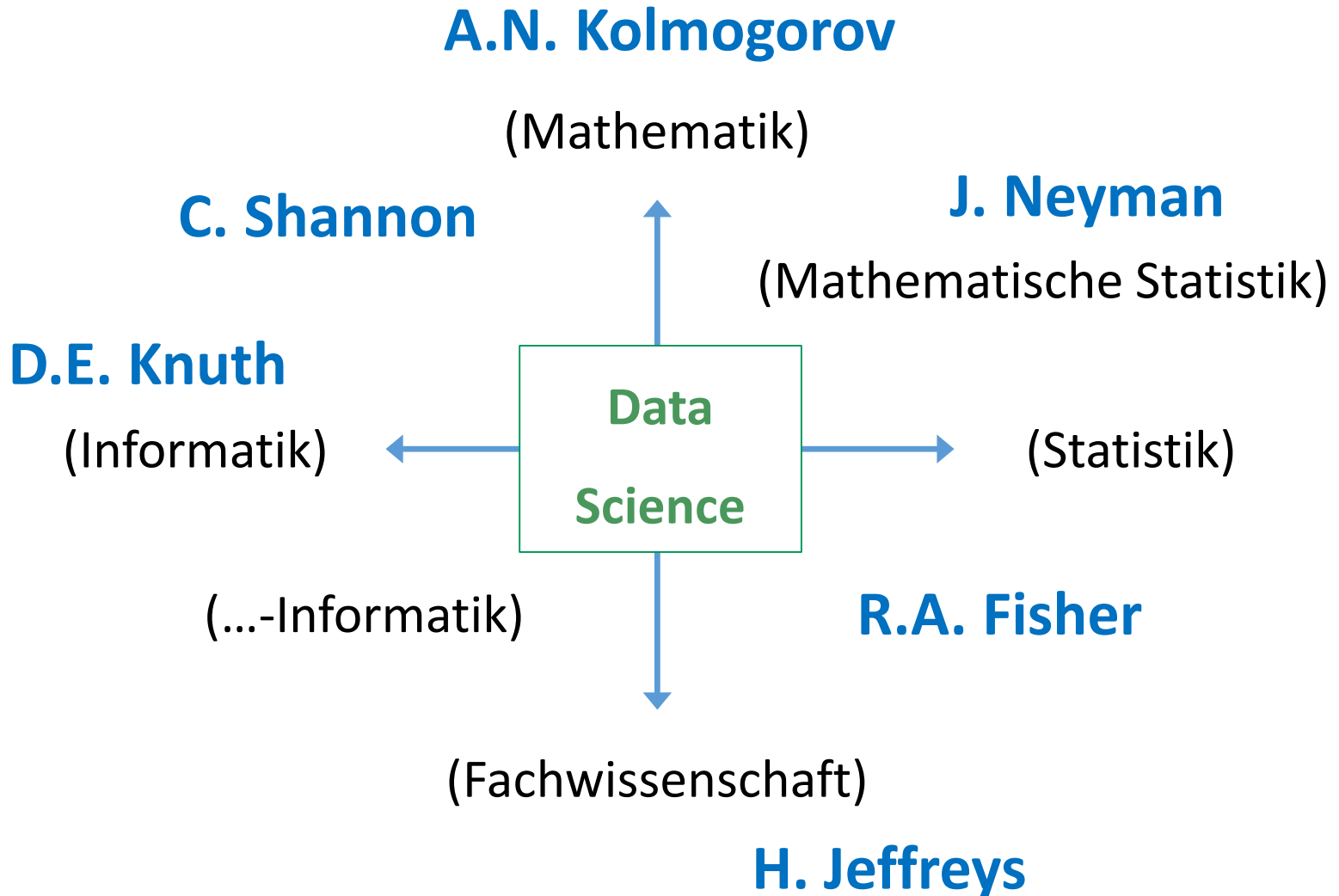




Warum Statistik nicht nur angewandte Mathematik ist

Uwe Saint-Mont

Verortung



Wer ist am Wichtigsten?

R.A. Fisher (1890-1962)

- 6 direkte Schüler, bis vor einem Monat
1288 akademische Nachfahren

J. Neyman (1894-1981)

- 43 direkte Schüler, bis vor einem Monat
2792 akademische Nachfahren

Quelle: Math. Genealogy Project

Wer ist am Wichtigsten?

H. Jeffreys (1891-1989)

- **8 direkte Schüler, bis vor einem Monat 887 akademische Nachfahren**

N. Kolmogorov (1903-1987)

- **82 direkte Schüler (insbesondere Gelfand, Dynkin, Sinai, Shiryaev), bis vor einem Monat 3251 akademische Nachfahren**

Daten als Fundament

Wissenschaftsbasierte Philosophie

Empirische Wissenschaft(en)

Statistik

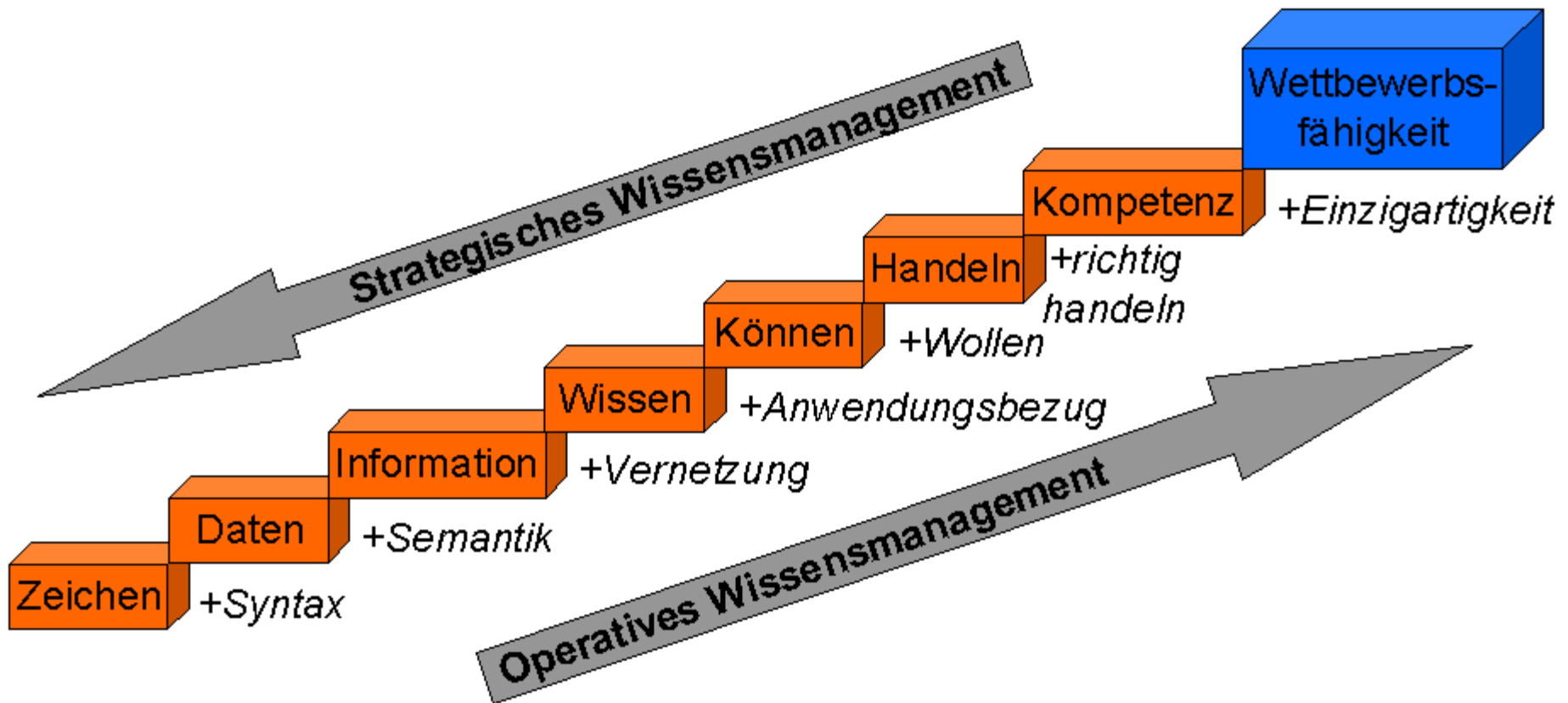
Informatik

Beobachtete Realität (Daten)

Daten als Fundament

- **Daten sind Zahlen in bzw. mit einem empirischen Kontext**
- **Daten sind der Rohstoff, aus dem Information und Wissen werden kann**

Data Science und BWL



„Wissenstreppe“ (K. North, 1998)

Wissenschaft von den Daten

- **Die Daten – und nichts anderes – stellen den Kontakt zur realen Welt / zum untersuchten Gegenstand / zur Empirie her und sind deshalb von entscheidender Bedeutung.**
- **Statistik ist die Wissenschaft, die sich mit der Erhebung, Analyse und Interpretation von Daten beschäftigt**

Priorität der klassischen Statistik: Daten-Erhebung

- **Einsicht: Fehler bei der Datenerhebung / Datensammlung können durch keine noch so „findige“ Art der Datenauswertung nachträglich wieder wettgemacht werden!**
- **„To consult a statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.“ R.A. Fisher (1938)**

Prioritäten (Menges, 1982)

- 1. Am wichtigsten sind die Daten**
- 2. Dann kommt die fachliche / inhaltliche Theorie und Interpretation**
- 3. Und dann erst die Analyse-Methoden**
Die Bedeutung der Methoden wird oftmals überschätzt!

Forschungsprozess

Realität

„beobachten“
„erheben“

„interpretieren“
„schlussfolgern“

(Zirkel)

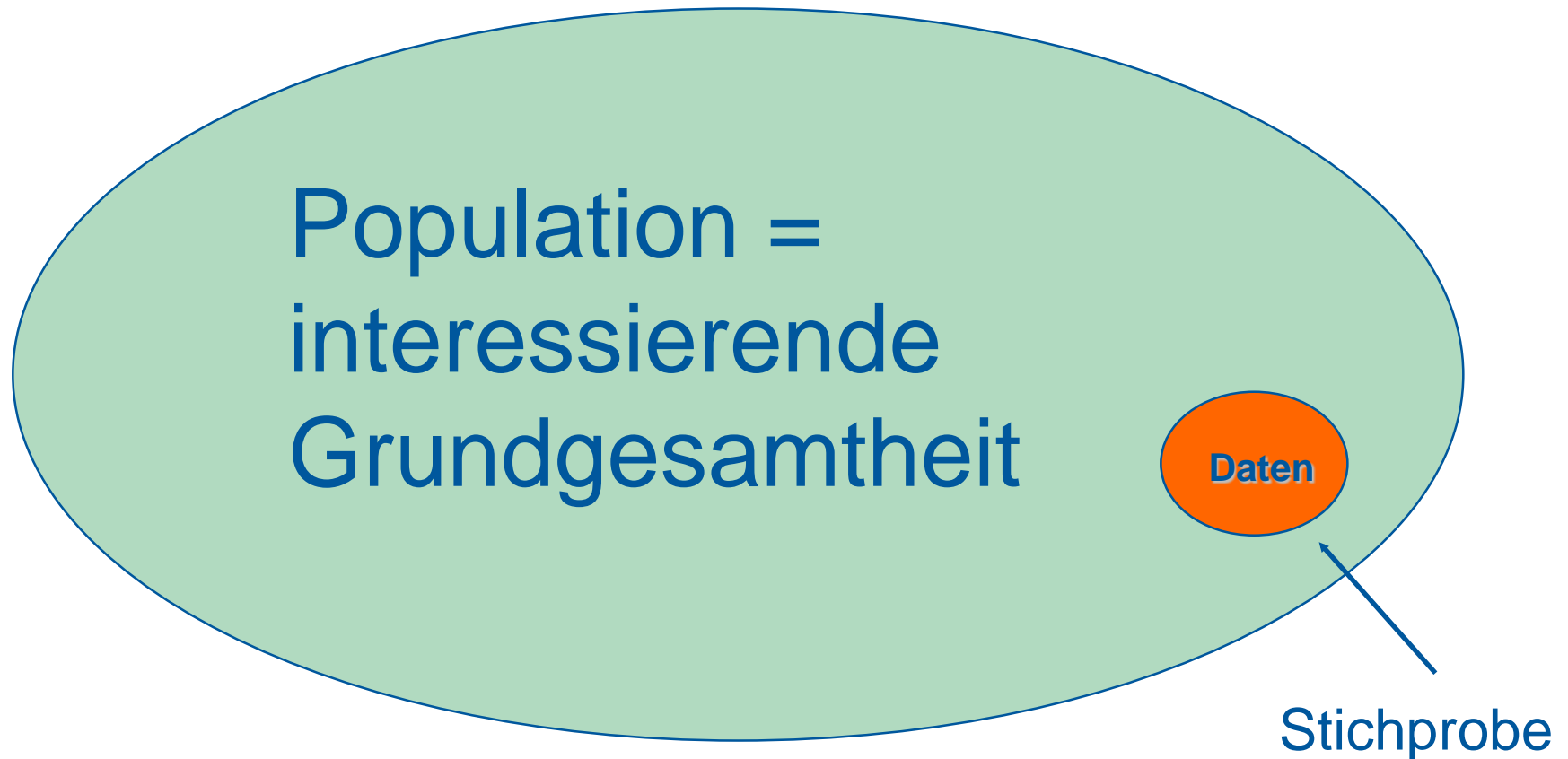
Ausschnitt
(Daten)

„auswerten“
„analysieren“

Kennwerte,
Grafiken,
Gruppierungen,
Modellierung

Eigenständigkeit:

- **Statistik ist nicht nur angewandte Mathematik und auch**
- **keine „Hilfswissenschaft“ zur Datenanalyse**
- **Sie hat viel mit Physik gemeinsam:**
 - empirische Wissenschaft
 - formal-mathematischer Kern
 - viele Verbindungen zu Nachbardisziplinen



„Pars pro toto“

Basismodell: zwei Ebenen

Population

Allgemein (Gesetzmäßigkeit)

Deduktion

Induktion

Stichprobe

Speziell (Daten)

Basismodell: zwei Ebenen (Bayes-Variante)

Posteriori-Wissensstand (Verteilung)

Information in den Daten



Priori-Wissensstand (Verteilung)

[objektiv, subjektiv, keine Information]

Basismodell: zwei Ebenen (nach Tukey, 1915-2000)

Allgemein (Gesetzmäßigkeit)

Konfirmatorische
Datenanalyse

Explorative
Datenanalyse

Speziell (Daten)

Bücher R.A. Fishers:

- **Statistical methods for research workers (1925)**
- **The design of experiments (1935)**
- **Statistical methods and scientific inference (1956)**

Basismodell: zwei Ebenen (nach Fisher)

Population



Experimentelles Design

Stichprobe

R.A. Fisher (1935)

- It is possible, and indeed it is all too frequent, for an experiment to be so conducted that no valid estimate of error is available. In such a case the experiment cannot be said, strictly, to be capable of proving anything.
- Perhaps it should not, in this case, be called an *experiment* at all, but be added merely to the body of *experience* on which, for lack of anything better, we may have to base our opinions.

R.A. Fisher (1935)

empfiehlt Randomized Controlled Trials (RCTs)

- Orientieren sich am naturwissenschaftlichen, kontrollierten, replizierbaren Experiment
- Zufällige Zuordnung von Probanden auf mehrere Gruppen (valide Statistik, Vergleichbarkeit), denn
- “[Randomization] relieves the experimenter from the anxiety of considering and estimating the magnitude of the innumerable causes by which the data may be disturbed.”

Basismodell: zwei Ebenen (nach Fisher)

Population

Allgemein (Gesetzmäßigkeit)

[starkes]

[weitreichende]

Design

Schlussfolgerung

Stichprobe

Speziell (Daten)

R.A. Fisher (1935)

zum allgemeinen Induktionsproblem:

- **We may at once admit that any inference from the particular to the general must be attended with some degree of uncertainty,**
- **but this is not the same as to admit that such inference cannot be absolutely rigorous,**
- **for the nature and degree of the uncertainty may itself be capable of rigorous expression.**

Die Rolle der Mathematik

sollte nicht unterschätzt werden! Sie sorgt für:

- **Präzision „in allen Teilen“
Problemmixierung, Argumentation,...**
- **logische (Folge-)Richtigkeit**
- **inneren Zusammenhalt**
- **eine Vielzahl von Verfahren, Algorithmen,...**
- **quantitative Aussagen, Modelle, Prognosen**

Rolle der Mathematik

Längerfristig: „Mathematisierung“ vieler bislang eher vager Begriffe und Gebiete

- **Entscheidungen (A. Wald)**
- **Konflikte und Kooperation (Spieltheorie: Morgenstern, von Neumann, Nash,...)**
- **Optimierung, „Operations Research“**
- **Information (Shannon, Kolmogorov)**

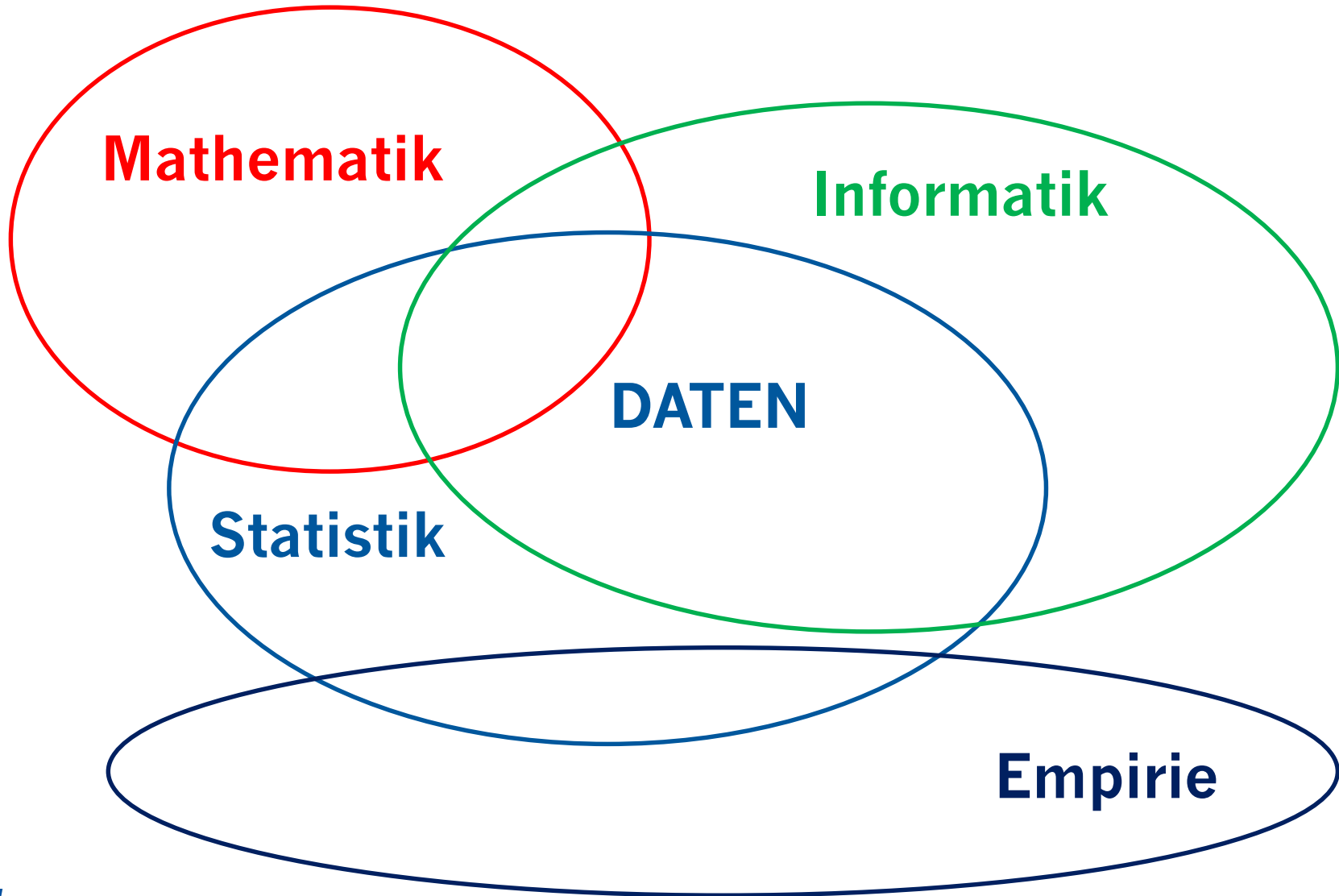
Rolle der Statistik

dringt nach und nach in Gebiete vor, wo sich Fehler (Streuung, Unschärfe, Unsicherheit, Risiken, Störgrößen...) nicht beliebig kontrollieren oder eliminieren lassen.

Statistics [is] the study of (Fisher 1925):

- **populations,**
- **variation,**
- **methods of the reduction of data**

Synthese: Data Science(s)



Literatur

- **Pyramide des Wissens: Saint-Mont, U. (2013).**
Die Macht der Daten, SpringerSpektrum.
- **“Forschungszirkel”:** Saint-Mont, U. (2011).
Statistik im Forschungsprozess, Physica.
- **Induktionsproblem: Saint-Mont, U. (2020).**
Induction: a logical analysis. Foundations of Science.
<https://doi.org/10.1007/s10699-020-09683-z>
- **Geschichte der Stochastik: Fischer, H. (Hrsg.)**
Springer, 2 Bde., ca. 2021/2022.