

Big Data: methodische und institutionelle Herausforderungen für die Soziologie

Rainer Diaz-Bone

(Universität Luzern)

„Big Data“

- Buzz word, Bedeutung unscharf (V-Definitionen wie Volume, Variety, Velocity ...) -> vorher Bezeichnungen wie „Data mining“
-> bereits seit den 1980er Jahren „Wirtschaftsinformatik“
- Dennoch grundlegende Veränderung der sozialen Welt: Sensoren, Digitalisierung, Datafizierung, Vernetzung, „Internet of Things“, „Industrie 4.0“
- Prozesse (z.B. in Unternehmen) und Lebenswelten (Haushalte, Märkte) sind zunehmend selbst Datenströme; Daten als das „neue Öl“ der Ökonomie

Problem: schon seit einigen Jahrzehnten findet digitale Transformation in Privatwirtschaft/Internetwirtschaft statt, nur ein Teil dieser Realität ist öffentlich „zugänglich“ (Twitterdaten, HTML-Seiten via Scraping)

„Big Data“

- Bislang

Sozialforschung: „Daten bilden soziale Realität ab“

Daten werden hier eigens erhoben, Datenformen, deren Verfügbarkeit und Bedeutung wird durch Forschende und Sozialforschungsinfrastrukturen definiert

- Neu

Big Data: „Soziale Realität ist datenförmig“ („Datenwelten“)

Menge, Verschiedenheit und hohe Auflösung von Daten ist „da“, Datenformen, deren Verfügbarkeit und Bedeutung wird durch andere Akteure (meist Unternehmen) definiert

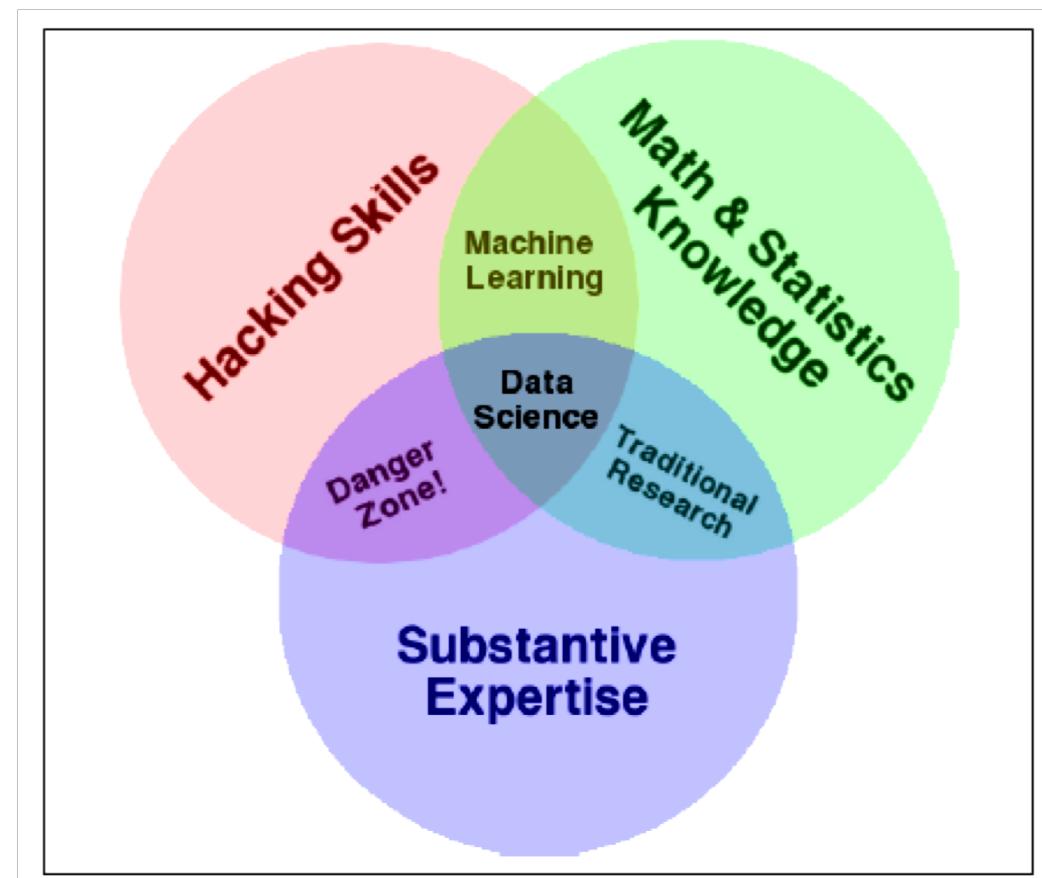
Kompetenzen

- Big Data meint nicht einfach nur „viel“, sondern heterogene und in der Regel wesentlich „maschinell“ generierte Daten, die heterogen und variabel anfallen
-> einlesen anspruchsvoll, zusammenbringen/organisieren, bereinigen/filtern
- Daten und Datenformate, die häufig nicht durch Sozialforschende generiert und (operationalisiert) sind, sondern in digitalen Infrastrukturen „anfallen“, danach ex post in analytische Formen und Kategorien zu bringen sind
-> Datenzugriffe programmieren, ex post kategorisieren/kodieren, für die Analyse vorbereiten
- Die Daten sind häufig Resultat einer Verkettung von menschlichen, organisatorischen und algorithmischen und technischen Abläufen
-> Datenprozesse sowie Kohärenz, Sinn und Qualität verstehen

Kompetenzen

- Keine ex ante Kontrolle über Operationalisierung, sondern ex post Bearbeitung/Exploration/Strukturierung der Daten
- Neue Kompetenzen erforderlich, was „Handling“ von Daten angeht
- Kenntnisse der technischen Plattformen und Schnittstellen (APIs) erforderlich
- Andere Softwareprogramme (wie R, Python, Datenbanken), die anderer Logik folgen als etablierte soziologische Software (Stata, IBM SPSS, SAS): „programmieren“ statt „Menüs klicken“
- Wichtig: Kompetenzen für die arbeitsteilige Kooperation mit IT-ExpertInnen (Plattformen und Programmierungen), StatistikerInnen (dort neue Ansätze wie AI), ExpertInnen für das untersuchte Feld

Kompetenzen



Quelle: Rahel Schutt & Cathy O'Neil (2014):
Doing data science, O'Reilly, S. 7

Kompetenzen

- Folgen:

Neue Bedeutung der Methodologie explorativer Datenanalyse und strukturentdeckender multivariater Verfahren, AI, Vorhersagemodelle; Folgen für „herkömmliche“ sozialwissenschaftliche Forschungslogiken?

Exploratives Vorgehen: Abduktives Schließen wird wichtiger, Modellierung und Strukturierung gewinnt an Bedeutung gegenüber deduktiver Prüfung von Hypothesen; interaktive Entwicklung von „Theorie“, „Befunden“

Daten: deren Sinn, Interpretation und Verständnis als Problem mit neuer Größenordnung (fehlendes Wissen über Einflüsse für Datenproduktion, statistische Ketten können intransparent sein, Inkohärenz verschiedener Technologien und Datenquellen)

Etablierte Konzepte wie Grundgesamtheit/Repräsentativität, Signifikanztests stoßen an Grenzen

Institutionelle Herausforderungen

- Situation Ausbildung:

Kaum existierende Lehrbücher, Methoden- und Statistikausbildung benennt das Phänomen Big data, aber ist noch an Primärerhebung und klassischer Sekundärdatenanalyse „reckeckiger“ Surveydatensätze orientiert

Studierende in den Sozialwissenschaften sind eher schwach in MINT-Kompetenzen, bereits die existierende Statistik- und Softwareausbildung überfordert viele Studierende

Lehrende auch in der Methodenausbildung verfügen nur selten in ausreichendem Umfang über die benannten Kompetenzen, derzeit bahnt sich dies eher auf Ebene der Qualifizierenden („Mittelbau“) an

Geeignete Lehrformen (praktische Seminare, Forschungsseminare) und insbesondere technologische Ausstattungen (eigene Server, Speicherkapazitäten) fehlen noch

Institutionelle Herausforderungen

- Marginalisierung der Soziologie?

In Wirklichkeit sind die effektiven Big data-Analysen seit Jahren bei den Unternehmen im Gang, dort findet auch Sozialforschung statt

- > unsichtbar für die Öffentlichkeit/Wissenschaft
- > Daten werden als ökonomische Ressource aufgefasst, hier gibt es Geschäftsmodelle wie Entwicklung von Vorhersagemodellen für Verhalten, Beeinflussung von (Kauf/Wahl)Entscheidungen durch zielgenaues Marketing etc.
- > technische Infrastrukturen und Daten werden durch überwiegend Private besessen und kontrolliert

Zwei Folgen: Asymmetrie in der Datenkontrolle und in vielen Bereichen absehbare Marginalisierung der Sozialwissenschaften für gesellschaftlich relevante Analysen

Institutionelle Herausforderungen

- „Bringing sociology back in“:

Was sind Formen für Kooperationen mit Privaten, die rechtlich, ethisch, wissenschaftlich vertretbar bzw. fruchtbar sind, insbesondere hier finden sich die vernetzten Devices/Sensoren, die in Echtzeit umfassend Verhaltensdaten erfassen

Wie bringt man Unternehmen dazu, Daten zu teilen?

-> Regulierung von privaten Dateninfrastrukturen

Kooperation mit den verstreuten öffentlich-rechtlichen Institutionen, die Daten generieren (BfS, aber auch Verbände, Versorger, Versicherer, Staat)

-> Etablierung von Dateninfrastrukturen (\neq Datensatzarchive)

Faktisch kann das Fach nur durch Kooperation mit IT-Menschen und interdisziplinär „Big data“ erschließen; zudem müssen Kompetenzen und Technologien in den Universitäten etabliert werden

Institutionelle Herausforderungen

- „Bringing sociology back in“:

(1) Soziologie muss Expertise entwickeln, wie man die vielen, technisch zum Teil sehr avancierten „Big Data“-Methoden hinsichtlich der sinnvollen soziologischen Verwendungen beurteilen lernt und diese dann adaptiert/einführt/anwendet

(2) Sinnfragen:

- was sind substantielle soziologische Theorieperspektiven, Fragestellungen für Big Data-Analysen? (Oder: „Big data and small questions/outputs“)
- wie erreicht man Kohärenz und Kontrolle/Verständnis von Big data (Datenströmen, statistischen Ketten), so dass Daten wissenschaftlich interpretierbar sind?
- wo bringen Algorithmen Grenzen der Kohärenz, Transparenz und Verstehbarkeit für Analyse ein?

(3) Soziologie muss sich als Disziplin verstehen, die Big data, dessen Antriebe sowie die dadurch ausgelösten sozialen Veränderungen, sozialen Probleme und ermöglichten gesellschaftlichen Chancen als Forschungsgegenstand annimmt

Institutionelle Herausforderungen

- Spezifikum der Soziologie:

Pluralität der Zugänge zu Big data:

- (1) methodisch: Das schillernde Phänomen „Big data“ stellt *neue* Technologien und methodologische Zugänge zur Verfügung
- (2) epistemologisch: Möglichkeit der Kritik und Kombination mit etablierten Methoden
- (3) soziologisch: Big data verändert Gesellschaft grundlegend, damit sind auch alle Bindestrich-Soziologien mit Digitalisierung der Gesellschaft konfrontiert
- (4) wissenschaftspolitisch: Big data hat Folgen für die gesellschaftliche Rolle/Position der Soziologie/der Sozialwissenschaften