



Die Geschichte von *Big Data*

Von Elena Aronova, Christine von Oertzen und David Sepkoski

Die zunehmende Allgegenwart riesiger Datenbanken und die schnelle technologische Weiterentwicklung von Datenspeicherung und -analyse hat vielerorts zu der Einschätzung geführt, die Wissenschaft des 21. Jahrhunderts sei in eine neue Ära eingetreten. Gleichzeitig mehren sich Stimmen, die auf historische Kontinuitäten von Praktiken und Epistemologien der Datenerzeugung und -analyse verweisen, die mehrere hundert Jahre zurückreichen. Die neue, von Elena

Aronova, Christine von Oertzen und David Sepkoski geleitete internationale Arbeitsgruppe strebt an, gemeinsam eine erste historisch breit angelegte Geschichte der wissenschaftlichen Erzeugung und des Umgangs mit Daten vorzulegen, die sowohl Datenpraktiken als auch deren Epistemologien, materielle Kultur und politische Konsequenzen über Disziplingrenzen hinweg vergleichend untersucht. Das kollektive Vorhaben will der aktuellen Debatte über das revolutionäre Potential datengesteuerter Methoden der Wissensgenerierung historische Tiefe verleihen. Die folgenden Fallstudien präsentieren drei Hauptarbeitsbereiche der Gruppe.



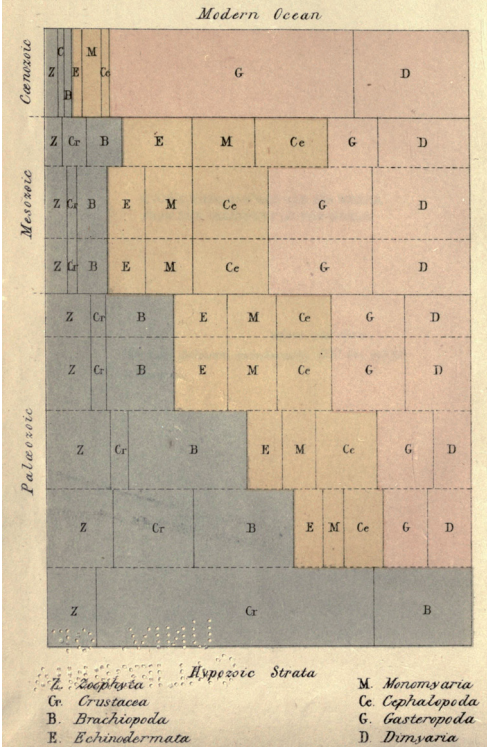
Typistin an einer Hollerith Stanztastatur im U.S. Census Bureau, 1908. Library of Congress Prints and Photographs Division Washington, D.C.

Die Veränderung von Datenpraktiken

Wie sind Daten und Praktiken ihrer Generierung und Analyse von Papiertechnologien über mechanische zu computerbasierten Informationstechnologien migriert? David Sepkoskis Projekt über „Die Datenbank vor dem Computer“ ist eine vergleichende Analyse der Datenverarbeitung in der Paläontologie und der Zoologie des 19. und 20. Jahrhunderts. Am Beispiel

SUCCESSIVE SYSTEMS OF MARINE
INVERTEBRAL LIFE.

Fig. 6



Frontispiz der „relativen Proportionen mehrerer Kategorien aufeinanderfolgender geologischer Perioden“, aus: John Phillips, *Life on the Earth; Its Origin and Succession*(1860).

groß angelegter Datensammlungen und der Auswertung von gedruckten taxonomischen Kompendien und schließlich digitalen Datenbanken führt Sepkoski vor, inwiefern die Forscher in diesen Feldern bereits im 19. Jahrhundert Datenbanken anlegten – lange bevor Computer aufkamen. Sepkoski widmet sich aber auch der Frage, inwieweit der Wandel von Technologien und materieller Kultur seit dem späten 19. Jahrhundert Datenpraktiken und Erkennt-

nisinteressen beeinflusst hat. Er zeigt, dass jede Epoche ihre eigenen Herausforderungen mit „Datenfriktionen“ zu meistern hatte, und neue Technologien und Materialitäten Folgen für die professionelle Arbeitsteilung in den taxonomischen Wissenschaften mit sich brachten.

Datenarbeit

Wie veränderten sich Arbeit und Technologie der Datenproduktion im historischen Verlauf? Christine von Oertzens Beitrag über „Die Mechanisierung der Zensusstatistik in Europa, 1890–1930“ untersucht die sozialen und politischen Bedingungen und Auswirkungen des Übergangs von der manuellen zur mechanischen Datenverarbeitung, zunächst vor allem bei Volkszählungen, aber auch in der Industrie und schließlich in der Kriegsführung. Als sich die amerikanische Erfindung der elektronischen Datenauszahlung 1890 unter europäischen Statistikern herumsprach, fanden die Maschinen unter den Experten der verschiedenen Länder sehr unterschiedliches Echo. Von Oertzen erklärt, warum Maschinenlesbarkeit in Russland und Österreich starke Befürworter fand, während britische und deutsche Statistiker davon Abstand nahmen, Maschinen für die Verarbeitung von Volkszählungsdaten einzusetzen. Im Deutschen Reich erzwang der Erste Weltkrieg mit seinen beispiellosen logistischen Herausforderungen ein Umdenken. Das Projekt zeigt, in welcher Weise der erste totale Krieg der Welt die Einführung der maschinellen Datenverarbeitung in Mitteleuropa beschleunigte.

Die Macht der Großdaten

Woher kommt die Macht von Großdaten? Elena Aronova erforscht in ihrem Projekt „Sind Großdaten politisch? Der Kalte Krieg und die politi-



Farbzeichnung von *aurora borealis* von William Crowder, *National Geographic* (1947).

sche Ökonomie des Datenaustauschs“, inwiefern Großdaten als Teil von Großforschung während des Kalten Kriegs neue Legitimität erhielten. Der Begriff der Großforschung (Big Science) kam in den frühen 1960er Jahren auf, um eine Organisation von Wissenschaft zu beschreiben, die im Zweiten Weltkrieg durch Unternehmungen wie dem Manhattan Project entstanden war und Weltraumstationen, die ersten Computer und Teilchenbeschleuniger umfasste: große Maschinen, die große Kosten verursachten und großes öffentliches Interesse fanden. Vor diesem Hintergrund rekonstruiert Aronova die Geschichte der Welt Datenzentren (World Data Centers), die im Rahmen des Internationalen Geophysischen Jahres (1957–1958) geschaffen

wurden. Aronova legt dar, dass die Praktiken und Technologien der Welt Datenzentren eng mit der politischen Ökonomie des Austauschs von Daten verknüpft waren und als solche zentrale Bedeutung für die Entwicklung der heutigen Datenbanken in den geophysischen Wissenschaften hatten.

Diese Beispiele zeigen, dass die jeweiligen konkreten Verbindungen zwischen Technologien, Praktiken, Materialitäten und Epistemologien ihre stets eigene Komplexität aufweisen. Während Technologien sich änderten, blieben Praktiken der Erzeugung und Verwaltung von Datenbanken und Instrumente der Organisation, Analyse und Repräsentation länger konstant.

b) Durch die Lochung zur maschinellen Zählung vorbereitete Lochkarte.

PERSONENSTATISTIK						HAUSHALTUNGSSTATISTIK										WOHNUNGSSTATISTIK							
Zählort		Geb. Jahr	Arbeitsort		Beruf	Ehlf.	Kind.	näh. Verw.		Gew. Geh.		Mieter		Dienstb.		and. Pers.	AKT F G	Wohn- räume	Küche	Schlaf- räume	Schl. St.	S. B.	
OA	Gem.		OA	Gem.				m	w	m	w	m	w	m	w								m
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9

Württembergische Volkszählung
 1910.
 Ist Wohnungsbervölkerung?

JA

Erste Lochkarte, die in einer deutschen Volkszählung in Einsatz kam, 1910, aus: Hermann Julius Losch, *Die Volkszählung vom 1. Dezember 1910* (1911), S. 184.

Computer haben die Verbreitung der Erscheinungsformen datengenerierter Forschung, die früheren materiellen Datenkulturen bereits inhärent waren, beschleunigt und verstärkt. Die Arbeitsgruppe hat zum Ziel, eine nuancierte Genealogie dieser Erscheinungsformen vorzulegen, die sowohl historische Kontinuitäten als auch genuine Brüche deutlich macht. Nur wenn diese Kontinuitäten und Brüche im historischen Kontext verständlich werden, können wir aktuelle Debatten über das revolutionierende Potential datenintensiver Formen der Wissensproduktion kritisch bereichern.

Elena Aronova ist seit 2012 wissenschaftliche Mitarbeiterin in Abteilung II (*Ideale und Praktiken der Rationalität*) am MPIWG (earonova@mpiwg-berlin.mpg.de).

Christine von Oertzen ist seit 2005 wissenschaftliche Mitarbeiterin in Abteilung II (*Ideale und Praktiken der Rationalität*) am MPIWG (coertzen@mpiwg-berlin.mpg.de).

David Sepkoski ist seit 2012 wissenschaftlicher Mitarbeiter in Abteilung II (*Ideale und Praktiken der Rationalität*) am MPIWG (dsepkoski@mpiwg-berlin.mpg.de).

Eine vollständige Version ist mit weiteren Forschungsthemen auf der Institutswebsite zugänglich („Aktuelles/Aktuelle Themen“).