



Historicizing Big Data

By Elena Aronova, Christine von Oertzen, and David Sepkoski

The increasing ubiquity of huge databases, coupled with rapid advances in technologies for storage and analysis of data, has suggested to some observers that twenty-first-century science has entered a new era. At the same time, however, scholars have begun to stress important historical continuities in data practices and epistemologies that stretch back over several hundred years. The new working group

led by Elena Aronova, Christine von Oertzen, and David Sepkoski aims to collectively present the first genuinely broad-scale history of practices, epistemologies, material culture, and political consequences of data across scientific disciplines, while adding much needed comparative dimensions and historical depth to the on-going discussion of the revolutionary potential of data-driven modes of knowledge production. The following case studies represent three main areas addressed by the working group.



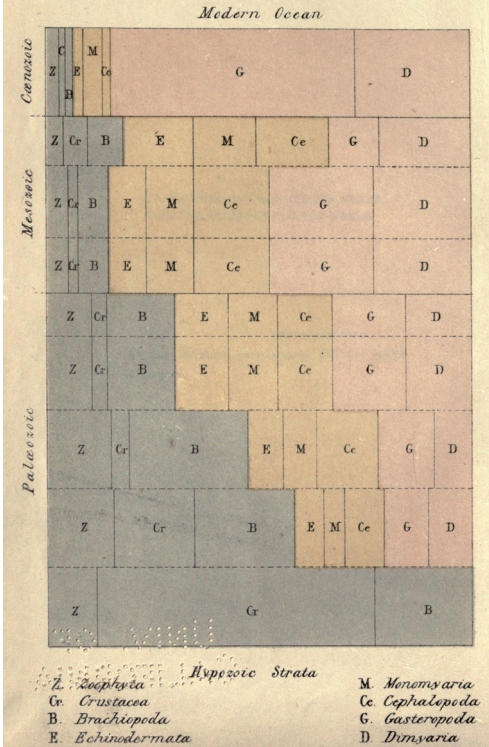
U.S. Census Bureau machines and operators, 1908. Library of Congress Prints and Photographs Division Washington, D.C.

The Transformation of Data Practices

How have data and data practices migrated from collection to paper-based then to mechanical and finally to computer-based information technologies? David Sepkoski's project on "The Database before the Computer" is a comparative analysis of nineteenth- and twentieth-century data approaches in paleontology and zoology. Using examples of large-

SUCCESSIVE SYSTEMS OF MARINE
INVERTEBRAL LIFE.

Fig. 6



Frontispiece, representing “the relative proportions of the several classes in successive geological periods.” John Phillips, *Life on the Earth; Its Origin and Succession* (1860).

scale quantitative data collection and analysis of both paper-based taxonomic compendia and eventual digital databases, he shows that practitioners in these fields were engaged in databasing long before computers arrived on the scene. However, he also attends to the ways that changes in the technology and material culture of data between the nineteenth and later twentieth centuries affected data practices and epistemologies, demonstrating that each era faced

particular challenges for coping with “data friction,” and new technologies and materialities of data had consequences for the professional division of labor in the taxonomic sciences.

Data Work

How did labor and technologies of data production change over time? Christine von Oertzen’s contribution on the “Mechanization of Census Statistics in Europe, 1890-1930” examines the social and political ramifications of the transition from manual to machine data processing across Europe between 1890 and 1930, first and foremost in census taking, but also in industry and warfare. When the word spread among European statisticians about the new American method of electronic tabulating in 1890, the machines appealed much more readily to some societies than others. Von Oertzen focuses on why machine readability found strong advocates in governments in Austria and Russia, whereas British as well as German census statisticians proved quite reluctant to introduce machines to the process of compiling census data. In the German Empire, the unprecedented challenges brought about by the First World War prompted an about-face toward big data machinery. The project will show on what terms the world’s first total war accelerated the adoption of tabulating machines in Central Europe.

The Power of Big Data

Where does the power of Big Data come from? Elena Aronova’s project “Do (Big) Data Have Politics? Cold War and the Political Economy of Data Exchange” explores how Big Data



Painting of *aurora borealis* by William Crowder, *National Geographic* (1947).

acquired renewed legitimacy during the Cold War era, as part of Big Science. Big Science as a notion was coined in early 1960s to describe the mode of organization of science originated in the Second World War and exemplified by such operations as the Manhattan project, space stations, early computers, and particle accelerators: accounts of big machines, big money, and big publicity that did not include big data. Against this background, Aronova's contribution draws on the history of the World Data Centers, organized to serve the International Geophysical Year (1957–1958), to show how the practices and technologies of the World Data Centers were intimately intertwined with the political economy of data exchange, and as such crucial for the develop-

ment of contemporary databases in the geophysical sciences.

As these examples show, the precise relationship between technologies, practices, materialities, and epistemologies of data is complex. While technologies have changed—from paper-based to mechanical to digital devices—database practices were more continuous than the technologies and tools used to organize, analyze and represent data. Computer technologies have accelerated and amplified features of data-driven science already present or latent in earlier material cultures of data. The working group aims to present a nuanced genealogy of these features of data-driven science that recognizes both underlying continuities, as

b) Durch die Lochung zur maschinellen Zählung vorbereitete Lochkarte.

PERSONENSTATISTIK										HAUSHALTUNGSSTATISTIK										WOHNUNGSSTATISTIK				
Zählort		Geb. Jahr		Arbeitsort		Einf.	Kind.	sbh. V-err		Gew. Geh.		Mieter		Dienstl.		and. Pers.	Größe	ANT F G	Wohnräume	Küche	Schlafräume	sonst. Z.B.		
ÖA	Gem.	v	n	x	x			m	w	m	w	m	w	m	w								m	w
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	M	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		
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	GN	3	3	3	3	3		
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	GD	4	4	4	4	4		
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	MD	5	5	5	5	5		
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	GMD	6	6	6	6	6		
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	Eig.	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		
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 Wohnung übervolker?

JA

First punch card used in a German census, 1910. From: Hermann Julius Losch, *Die Volkszählung vom 1. Dezember 1910* (1911), p. 184.

well as genuine ruptures. Only by considering these continuities and ruptures can we critically expand the on-going discussion of the revolutionary potential of data-intensive modes of knowledge production.

Elena Aronova is Research Scholar in Department II (*Ideals and Practices of Rationality*) at the MPIWG since 2012 (earonova@mpiwg-berlin.mpg.de).

Christine von Oertzen is Research Scholar in Department II (*Ideals and Practices of Rationality*) at the MPIWG since 2005 (coertzen@mpiwg-berlin.mpg.de).

David Sepkoski is Research Scholar in Department II (*Ideals and Practices of Rationality*) at the MPIWG since 2012 (dsepkoski@mpiwg-berlin.mpg.de).

The full version of this feature and more research topics are accessible at the Institute's website ("News/Feature Stories").