

## BIG DATA

# Intimate details

A long-forgotten social science archive offers a lesson in responsible data management

By Laura Stark

The U.S. Library of Congress holds a long-forgotten collection of intimate data, the fate of which Rebecca Lemov describes as “a parable for our time.” Lemov’s humane, hilarious, and smart new book, *Database of Dreams*, recovers the story of the enormous effort and expense that went into creating this analog database during the mid-20th century—only to have it fade from personal and institutional memory within decades. The book shows that, although some things are forgotten because they are unimportant, others lose importance because they are forgotten.

The database was the brainchild of Bert Kaplan, a Harvard Ph.D. and rising star in 1950s social science. His idea was to create an infrastructure to pool, preserve, and share “endangered data”: the raw, primary documents that individual social scientists had collected for their own research projects and were apt to throw away in the days when filing cabinets were state-of-the-art data storage systems. Big funders bankrolled the project, and at its peak, upwards of 100 research institutions worldwide had bought the data set.

When the project started in the 1950s, it was literally a database full of dreams, accounts of the nighttime figments inside people’s heads. Kaplan was among a cadre of anthropologists and psychologists who ventured out from universities to collect “fleeting thoughts, random asides, irreverent inquiries and sad memories, life stories and dreams” from people in the flux of modernization. Their aim was both patronizing and ambitious: to collect ephemeral data from communities that were likewise fading from modern life—living relics of vanishing worlds.

The database would eventually expand to include life histories and the results of Rorschach and other projective tests from members of a hodgepodge of communities around the world, including small-town Midwesterners, resettled American Indians, and natives of the Philippine Islands, to

name a few. It can be hard to appreciate the authority once accorded to data that today seem the mere quirk and color of history. Yet, projective tests were taken so seriously in the 1940s and 1950s that scientists used them to access the psychic depths of Nazi war criminals.

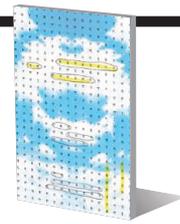
The archive was also a “database of dreams” in a second sense. For social scientists, it represented the aspiration to catalog the human condition. Such was the enthusiasm for the project that in only 7 years, 60 researchers freely donated their data to be included in the archive.

Within decades, the database faded from use and memory, in part because of Kaplan’s unlucky choice of platform: the Microcard. But there was more than platform at play, and to her credit, Lemov avoids the simplistic view that the commercial failure of the Microcard explains everything.

The great virtue, and fatal flaw, of the database was that it stored raw, uninterpreted intimate material that users could then enroll in the service of their own claims. In 1964, a *Newsweek* journalist did just this, imposing his own, politically provocative interpretation on data Kaplan had collected. The debacle enraged the communities

## Database of Dreams The Lost Quest to Catalog Humanity

Rebecca Lemov  
Yale University Press,  
2015. 377 pp.



Kaplan studied, turned researchers against him, and distressed Kaplan himself. This event alone did not bring the demise of the database (although it certainly altered Kaplan’s career). Yet, it was emblematic of the pervasive worry that users might fail to exercise restraint and good judgment with the database.

Today, many scholars are concerned about big data and the consequences of

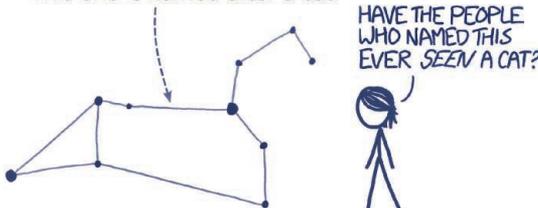
“biocapital,” the new financial assets made possible through bioinformatics. Lemov, however, is interested in the responsibilities attached to intimate data in a world in which the long-term effects of personal revelations are impossible to predict.

The fundamental question that Lemov pursues is how the database of dreams can come to seem like a thing of the past. After all, the database still exists in the here and now: in the Library of Congress, in research libraries across the world, and even on Lemov’s hard drive. It is the question of how to respond to the paradox of intimate data and is, Lemov writes, “a symptom of our own future buried in the present.”

10.1126/science.aad2423

**“The book shows that, although some things are forgotten because they are unimportant, others lose importance because they are forgotten.”**

This one is named after a cat.



In *Thing Explainer*, xkcd creator Randall Munroe sets out to demystify a wide range of complex systems and natural phenomena using only the thousand most common words in the English language. The book features simple blueprints annotated with short descriptions that explain everything from the layout of the “shared space house” (international space station) to the causes of “great circle storms” (hurricanes) to the science behind the “food-heating radio box” (microwave). Munroe’s signature humor and firm grasp on the underlying science and engineering make the book a delightful and informative read.

## POPULAR SCIENCE

### Thing Explainer Complicated Stuff in Simple Words

Randall Munroe  
Houghton Mifflin  
Harcourt, 2015. 72 pp.



10.1126/science.aad7700

The reviewer is at the Center for Medicine, Health, and Society, Vanderbilt University, Nashville, TN 37235, USA. E-mail: laura.stark@vanderbilt.edu