In 1969, Monsanto executives held a top-secret meeting. Public anxiety was mounting over the human health hazards of polychlorinated biphenyls (PCBs), one of the company’s top revenue streams. "We could go out of business," says one of the suits. Or, says another, "Sell the hell out of them."

Memorialized in handwritten minutes buried for decades in company archives and made public through legal proceedings, details are online at ToxicDocs.org.

The world's largest searchable database of internal corporate documents on industrial pollutants, ToxicDocs was featured in a special issue of the Journal of Public Health Policy. "This enormous stash of material represents the unvarnished opinions and inside dealings of the companies that make some of the most toxic products," says project principal David Rosner, PhD, MPH, the Ronald H. Lauterstein Professor and co-director of the Center for History and Ethics of Public Health at the Columbia Mailman School, which developed and maintains the site. "Corporate archives are curated by the companies to cast themselves in a favorable light. ToxicDocs tells the whole story."

Making sense of the 20 million pages that comprise ToxicDocs was like "searching for a needle in a haystack," says Merlin Chowkwanyun, the assistant professor of Sociomedical Sciences who led the effort to digitize and index the material, which is now machine-readable and searchable by keyword. The School's investment in high-performance computing—the art of pooling thousands of servers and applying their collective power to ambitious tasks—was critical to the effort, says Chowkwanyun.

Additional investigative tools are still in development, with the support of a three-year grant from the National Science Foundation. Among the new tools, one will allow users to identify patterns, such as common phrases and recurring names of people and entities, across clusters of documents. This "relationship miner" will also feature a graphical representation that illustrates the strength of connections between keywords. Other tools will distinguish among document types—memos exchanged among executives, unpublished scientific studies, public relations campaign plans, letters to policymakers, and classified meeting minutes, for example—and show the rise and fall of certain terms over time.

"A single document by itself doesn't tell the whole story," says Chowkwanyun. "ToxicDocs connects the dots. This larger data set paints a much bigger picture."