



The Journal of Civic Information

Volume 3 | Number 1

June 2021

Journal homepage: <https://journals.flvc.org/civic/>

ISSN (online): 2641-970X

Freedom of the Database: Auditing Access to Structured Data

Jonathan Anderson and Sarah K. Wiley *

Article Information

Received: March 13, 2021

Accepted: April 14, 2021

Published: June 30, 2021

Keywords

Freedom of information
FOIA
Databases
Data journalism
Public records

Abstract

As government operations at all levels have become increasingly computerized, records of those activities have moved from paper to databases. Yet there has been little empirical research about the public's ability to access such records in practice. This study uses field research to assess how 44 public universities respond to records requests of varying complexity for structured data. Sampled universities produced responsive structured data without a fee in slightly more than a quarter of requests, meaning the vast majority of requests failed to yield the information sought in a structured format and for free.

* Jonathan Anderson is a doctoral student at the Hubbard School of Journalism and Mass Communication at the University of Minnesota. Sarah K. Wiley, J.D., is a doctoral candidate at the Hubbard School of Journalism and Mass Communication at the University of Minnesota. Please send correspondence about this article to Jonathan Anderson at and08164@umn.edu. An earlier version of this paper was presented at the National Freedom of Information Coalition summit FOI Research Competition Sept. 10, 2020.

To cite this article in Bluebook: Jonathan Anderson and Sarah K. Wiley, *Freedom of the Database: Auditing Access to Structured Data*, 3(1) J. CIVIC INFO 30 (2021).

To cite this article in APA: Anderson, J., & Wiley, S. K. (2020). Freedom of the database: Auditing access to structured data. *Journal of Civic Information*, 3(1), 30-56.

DOI: 10.32473/joci.v3i1.129180

Published under Creative Commons License CC BY-NC, Attribution NonCommercial 4.0 International.

Freedom of the database: Auditing access to structured data

In 2019, when ProPublica journalists set out to investigate financial ties between medical professors and healthcare companies, they filed records requests with public universities around the United States for faculty conflict-of-interest and outside income information.¹ Getting the records proved challenging.² Nearly 30 public universities denied the requests or attempted to charge onerous fees, according to the news organization.³ For instance, the New York Joint Commission on Public Ethics refused to release a database of financial disclosure data, arguing that it “would be a prohibitive task for the Commission given the very small number of staff members who work to fulfill the record requests we receive.”⁴ Likewise, the University of Alaska told ProPublica that it could pull from a database only the names of faculty who filed conflict-of-interest forms but not details on how much outside money faculty collected or from where.⁵ “The data fields you mentioned like title, outside interest, type of activity and compensation type of information are not stored in the database and is not available during the extraction,” a University of Alaska data privacy and compliance officer told ProPublica.⁶ New York and Alaska instead offered to provide financial details about only those faculty whom reporters named specifically.⁷

Anecdotal accounts suggest ProPublica’s experience is not unique. Others have reported difficulty with accessing structured data⁸ because of purported technological limitations, processing costs, claims by the government that fulfilling data requests requires creation of a new record, and arguments against disclosure premised on the unique characteristics of big data.⁹ Yet these problems have received little empirical attention, a gap in knowledge that is particularly acute considering

¹ Annie Waldman, *Medical Professors Are Supposed to Share Their Outside Income With the University of California. But Many Don’t*, PROPUBLICA (Dec. 6, 2019), <https://www.propublica.org/article/medical-professors-are-supposed-to-share-their-outside-income-with-the-university-of-california-but-many-dont>.

² Annie Waldman & David Armstrong, *We Asked Public Universities for Their Professors’ Conflicts of Interest — and Got the Runaround*, PROPUBLICA (Dec. 6, 2019), <https://www.propublica.org/article/we-asked-public-universities-for-their-professors-conflicts-of-interest-and-got-the-runaround>.

³ *Id.* Such obstacles are not new. In 1978, a Wisconsin newspaper sued the University of Wisconsin-Madison for access to faculty outside income reports. The university asserted that disclosure would violate professors’ academic freedom, but a judge rejected that argument and ordered the university to release the reports. See David Pritchard & Jonathan Anderson, *Forty Years of Public Records Litigation Involving the University of Wisconsin: An Empirical Study*, 44 J.C. & U.L. 48 (2018-2019).

⁴ *Id.*

⁵ *Id.*

⁶ *Id.*

⁷ *Id.*

⁸ By “structured data” we mean clearly defined data types whose pattern makes them easily searchable, such as relational databases, comma-separated value (CSV) files, and Microsoft Excel spreadsheets.

⁹ Dep’t of Justice v. Reporters Comm. For Freedom of Press, 489 U.S. 749 (1989); Jennifer LaFleur, *Data and Transparency: Perils and Progress*, in TRANSPARENCY IN POL. & THE MEDIA: ACCOUNTABILITY AND OPEN GOV. (Nigel Bowles, James T. Hamilton, & David A. L. Levy, eds., 2014); Joey Senat, *Public Access and Informational Privacy in Electronic Government Databases*, in TRANSPARENCY 2.0: DIGITAL DATA AND PRIVACY IN A WIRED WORLD 36 (Charles N. Davis & David Cuillier, eds., 2014); Associated Press, *Freedom of Information, Access Issues in all 50 States*, STAR TRIB. (Minneapolis, Minn.) (Mar. 13, 2015), <https://www.startribune.com/freedom-of-information-access-issues-in-all-50-states/296168991/>; *Cost-Related Access Challenges, Solutions in 18 States*, ASSOCIATED PRESS (Mar. 13, 2015), <https://www.ap.org/ap-in-the-news/2015/cost-related-access-challenges-solutions-in-18-states>; D. Victoria Baranetsky, *Data Journalism and the Law*, COLUM. J. REV. (Sept. 19, 2018), https://www.cjr.org/tow_center_reports/data-journalism-and-the-law.php; Sarah Ryley, *What I Learned From Making Dozens of Public Records Requests for Police Data*, THE TRACE (Jan. 17, 2020), <https://www.thetrace.org/2020/01/police-data-documentation-public-records-requests/>; Sarah Kay Wiley, *The Grey Area: How Regulations Impact Autonomy in Computational Journalism*, DIGITAL JOURNALISM (Mar. 17, 2021).

government information has for decades been increasingly stored in computers and databases.¹⁰ It is thus critical to understand the level of access to such records, and what that may mean for the public's right to know.

This study uses field research to test how well state government agencies—in this case, 44 public universities—respond to public records requests of varying complexity for structured data. Importantly, the principal purpose of the study is not to assess what the law says about access to government data, although that is discussed to some extent. Rather, the study examines actual access. The article contributes to scholarly literature about how public record laws operate in practice, especially at the state and local government levels,¹¹ and builds on a growing body of academic research that has used freedom of information audits as the primary method of data collection.¹²

The paper is organized as follows. We first review pertinent literature about public records laws, access to government data, and the importance of data in journalism. Second, we situate the study within previous research about access to electronic government information and studies that have used access audits and similar tools to assess how public records laws operate in practice. After walking through the study's research questions and methods, we present our findings. We then discuss the implications of these findings and offer suggestions for journalists seeking data, governments responding to records requests for data, and lawmakers who could reform and update public records laws. We also identify several pathways for future research on access to structured data.

Literature review

Public records laws and the right to know

The legal right to access government-held records in the United States is principally derived from public records laws.¹³ Every state has its own version of a public records law, which provides a general right of access to records typically created or maintained by state and local governments.¹⁴

¹⁰ Sarah Giest & Reuben Ng, *Big Data Applications in Governance and Policy*, 6 POL. & GOVERNANCE (2018); Danah Boyd & Kate Crawford, *Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon*, 15 INFO., COMM. & SOC'Y 662 (2012); Matthew D. Bunker, Sigman L. Splichal, Bill Chamberlin, & Linda M. Perry, *Access to Government-Held Information in the Computer Age: Applying Legal Doctrine to Emerging Technology*, 20 FLA. ST. U. L. REV. 543 (1993); Leo T. Sorokin, *The Computerization of Government Information: Does It Circumvent Public Access under the Freedom of Information Act and the Depository Library Program*, 24 COLUM. J.L. & SOC. PROBS. 267 (1991).

¹¹ Richard J. Peltz-Steele & Robert Steinbuch, *Transparency Blind Spot: A Response to Transparency Deserts*, 48 RUTGERS L. REC. 1 (2020).

¹² See, e.g., David Cuillier, *Honey v. Vinegar: Testing Compliance-Gaining Theories in the Context of Freedom of Information Laws*, 3 COMM. L. & POL'Y 203 (2010); Katherine Fink, *Freedom of Information in Community Journalism*, 7 COMM. JOURNALISM 17 (2019); A. Jay Wagner, *Piercing the Veil: Examining Demographic and Political Variables in State FOI Law Administration*, 38 GOV'T INFO. Q. 1 (2020); Peter Spáč, Petr Voda, & Jozef Zagrpanc, *Does the Freedom of Information Law Increase Transparency at the Local Level? Evidence from a Field Experiment*, 35 GOV'T INFO. Q. 408 (2018); Stephan Grimmelikhuijsen, Peter John, Albert Meijer, & Ben Worthy, *Do Freedom of Information Laws Increase Transparency of Government? A Replication of a Field Experiment*, 1 J. OF BEHAVIORAL PUB. HEALTH ADMIN, 1 (2018); Ben Worthy, Peter John, & Matia Vannoni, *Transparency at the Parish Pump: A Field Experiment to Measure the Effectiveness of Freedom of Information Requests in England*, 27 J. OF PUB. ADMIN. RESEARCH & THEORY 485 (2017); James ben-Aaron, Matthew Denny, Bruce Desmarais, & Hanna Wallach, *Transparency by Conformity: A Field Experiment Evaluating Openness in Local Governments*, 77 PUB. ADMIN. REV., 68 (2017); Paul Lagunes & Oscar Pocasangre, *Dynamic Transparency: An Audit of Mexico's Freedom of Information Act*, 97 PUB. ADMIN. 162 (2019).

¹³ Christina Koningisor, *Transparency Deserts*, 114 NW. U. L. REV. 1461 (2020).

¹⁴ *Id.*

The federal government also has a public records law, the Freedom of Information Act, which has influenced the development of some state laws.¹⁵

Public records laws differ in significant ways. Most of the laws are statutory, but some are also rooted in state constitutional provisions.¹⁶ Laws also differ in terms of what types of entities have disclosure obligations, what information is subject to release and what information can be withheld, response time requirements, permissible fees, and dispute-resolution processes.¹⁷

Yet there are some similarities. All state public records laws generally define records subject to disclosure broadly—not limited to only those documents public officials are required to keep under state record-retention rules.¹⁸ Public records also are typically defined based on the information contained in them, regardless of the medium or format of the record.¹⁹ While some states allow agencies to ask requesters why they want information, requesters are generally not required to answer.²⁰ Every state provides an opportunity for requesters to obtain at least some records for free, either through explicit fee provisions granting free access or by making fees discretionary.²¹ Requesters also are entitled to obtain copies of records in every state and have a right to obtain independent review of a dispute.²²

Public records laws are premised on a patchwork of related rationales and policy justifications. One common basis for the right to know is that citizens are the ultimate decisionmakers in a representative democracy.²³ Harold Cross articulated this notion in 1953: “Public business is the public’s business. The people have the right to know,” Cross wrote.²⁴ “Freedom of information is their just heritage. Without that the citizens of a democracy have but changed their kings.” Legal theorist Thomas Emerson later argued that the “public, as sovereign, must have all information available in order to instruct its servants, the government.”²⁵

Another basis for public records laws is that voters need to be informed so they can intelligently perform their civic roles, such as voting for public candidates and evaluating the propriety of public policy. U.S. Rep. John Moss, who helped shepherd FOIA through Congress, articulated this justification on the House floor.²⁶ “Our system of government is based on the participation of the governed, and as our population grows in numbers it is essential that it also grow in knowledge and understanding,” Moss said.²⁷ “We must remove every barrier to information about—and understanding of—Government activities consistent with our security if the American public is to be adequately equipped to fulfill the ever more demanding role of responsible citizenship.”²⁸

¹⁵ Roger A. Nowadzky, *A Comparative Analysis of Public Records Statutes*, 28 URB. LAW. 65 (1996).

¹⁶ See, e.g., FLA. CONST. art. I, § 24; CAL. CONST. art. I, § 3.

¹⁷ Nowadzky, *supra* note 15.

¹⁸ Jonathan Anderson & David Pritchard, Policy Liberalism and Access to Information in the American States (Aug. 2020) (unpublished paper presented at the annual meeting of the Association for Education in Journalism and Mass Communication).

¹⁹ *Id.*

²⁰ *Id.*

²¹ *Id.*

²² *Id.*

²³ Thomas I. Emerson, *Legal Foundations of the Right to Know*, 1976 WASH. U. L. Q. 1-24, 17 (1976). See also Thomas I. Emerson, *Toward a General Theory of the First Amendment*, 72 Yale L.J. 877 (1962-1963).

²⁴ HAROLD L. CROSS, *THE PEOPLE’S RIGHT TO KNOW* (1953).

²⁵ *Id.*

²⁶ 112 Cong. Rec. 13007 (1966) (statement of Rep. John Moss).

²⁷ *Id.*

²⁸ *Id.*

Public records laws also are thought to help facilitate government accountability. In *Department of Air Force v. Rose*, 425 U.S. 352 (1976), the U.S. Supreme Court wrote that FOIA was intended to “pierce the veil of administrative secrecy and to open agency action to the light of public scrutiny.”²⁹ Legislatures have similarly invoked this ideal. For example, the Illinois Freedom of Information Act declares that access to public records is “necessary to enable the people to fulfill their duties of discussing public issues fully and freely, making informed political judgments and monitoring government to ensure that it is being conducted in the public interest.”³⁰

The rise of the government database

Public agencies at the federal, state, and local levels are increasingly collecting and using data.³¹ Mass transit systems log every swipe of a bus or train pass, schools track students’ attention levels,³² and criminal justice agencies use a number of data points to predict recidivism rates.³³ This massive amount of data is collected and preserved for a period of time within electronic databases. While government officials state that data governance and transparency have become a top priority,³⁴ open government advocates remain skeptical as both public awareness of and access to databases is often lacking.³⁵

Unlike the federal Privacy Act’s requirement that each federal agency that maintains a system of records must announce the existence of a database, no such requirement is uniformly mandated to all state and local agencies.³⁶ Rather, states (or in the absence of state direction, local agencies) decide whether to provide proactive notice of databases they maintain. California is one example. It enacted a law in 2015 requiring local governments to affirmatively disclose a catalog listing databases and other information systems they use, subject to certain exceptions.³⁷ The catalogs must also be posted online if local agencies have a website.³⁸ To illustrate, the city of Salinas, California, which has a population of about 156,000 residents, uses at least nine information systems to manage electronic data about permitting and land use, city finances, streets, and crimes, among other purposes.³⁹

Obtaining access to electronic databases has also been fraught with litigation at both the state⁴⁰ and federal level.⁴¹ For example, governments have argued that access to electronic databases

²⁹ *Dep’t of Air Force v. Rose*, 425 U.S. 352, 361 (1976).

³⁰ 5 ILL. COMP. STAT. ANN. 140/1.

³¹ Boyd & Crawford, *supra* note 10; Giest & Ng, *supra* note 10.

³² Juliane Jarke & Andreas Breiter, *Editorial: The Datafication of Education*, 44 LEARNING, MEDIA & TECH (2019); Erika Gimbel, *Biometric Tech Can Track How Well Students Are Paying Attention*, ED TECH (Feb. 23, 2018), <https://edtechmagazine.com/k12/article/2018/02/biometric-technology-tracks-students-attention>.

³³ Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System* 70 STAN. L. REV. 1343 (2018).

³⁴ Chief Information Officers Council, Letter on Federal Data Strategy, <https://www.cio.gov/policies-and-priorities/data/>.

³⁵ Anneke Zuiderwijk, et al., *Investigating the Attainment of Open Government Data Objectives: Is there a Mismatch Between Objectives and Results?* 85 INT’L. REV. OF ADMIN. SCI. 645, 672 (2019).

³⁶ See 5 U.S.C. § 552a(e)(4). These notices are commonly referred to as “systems of records notices” or “SORNs.”

³⁷ CAL. GOV’T CODE § 6270.5 (West). See also Sen. Bill No. 272 (2015-2016 Reg. Sess.), https://leginfo.ca.gov/faces/billNavClient.xhtml?bill_id=201520160SB272.

³⁸ *Id.*

³⁹ City of Salinas, California, Public Information Request, <https://www.cityofsalinas.org/our-city-services/city-clerk/public-information-request>.

⁴⁰ See e.g., *Courts Lock up Public Databases*, Reporters Comm. for Freedom of the Press, <https://www.rcfp.org/journals/courts-lock-public-database/>.

⁴¹ *Ctr. for Investigative Reporting v. Dep’t of Justice*, No. 17-CV-06557-JSC, 2018 WL 3368884 (N.D. Cal. July 10, 2018); *Ctr. for Investigative Reporting v. Dep’t of Justice*, 982 F.3d 668 (9th Cir. 2020); *Dep’t of Justice v. Reporters Comm. for Freedom of Press*, 489 U.S. 749 (1989).

hinders privacy;⁴² that pulling a subset of data is akin to creating a “new record” and thus not required;⁴³ that only custodians of a database may provide access;⁴⁴ and that providing a copy of data would violate the copyright of the software used in connection with creating and maintaining the database.⁴⁵ Such arguments highlight the need for updates to public record laws in order to facilitate government transparency and accountability as technology changes and modes of information management evolve.

Accountability journalism in the Age of Data

Data and journalism have also become increasingly intertwined,⁴⁶ so much so that several scholars have opined that journalism appears to be taking a “quantitative turn.”⁴⁷ Such observations seem to be backed by industry trends. For example, although the local newspaper industry has been plagued with layoffs and closures across the nation, several large nationally focused organizations, including *The Washington Post*,⁴⁸ *Wall Street Journal*,⁴⁹ and Bloomberg News⁵⁰ are increasing their use of computational tools and expanding their data journalism and engineering teams. Meanwhile, recent attendance at the annual National Institute for Computer-Assisted Reporting (NICAR) conference has hit record levels.

Journalists are increasingly applying computational methods to analyze and examine governmental data.⁵¹ Both scholars and journalists have illustrated how computational methods can help journalists generate story leads and find patterns across documents.⁵² For example, Broussard demonstrated how artificial intelligence could be used to flag important differences between school expenditures found in educational data.⁵³ Furthermore, researchers have also emphasized how the capability to apply computational methods is cost-effective, saving both time and money. As Cohen, Hamilton, and Turner explain: “Stories will emerge from stacks of financial disclosure forms, court records, legislative hearings, officials’ calendars or meeting notes, and regulators’ email messages

⁴² Dep’t of Justice v. Reporters Comm. for Freedom of Press, 489 U.S. 749 (1989).

⁴³ See e.g., Ctr. for Investigative Reporting v. Dep’t of Justice, No. 17-CV-06557-JSC, 2018 WL 3368884 (N.D. Cal. July 10, 2018); Ctr. for Investigative Reporting v. Dep’t of Justice, 982 F.3d 668 (9th Cir. 2020).

⁴⁴ LexisNexis Risk Data Mgmt. Inc. v. North Carolina Admin. Office of Courts, 754 S.E. 2d 223, 2015 WL 4999586 (N.C. August 21, 2015) (No. 101PA14).

⁴⁵ Pictometry v. FOIC, 307 Conn. 648 (2013); State ex rel. Gambill v. Opperman, 135 Ohio St.3d 298 (2013).

⁴⁶ Samantha Sunne, *The Rise of Data Reporting*, AMERICAN PRESS INSTITUTE (Mar. 9, 2019), <https://www.americanpressinstitute.org/publications/reports/strategy-studies/data-reporting-rise/>.

⁴⁷ Caitlin Petre, *A Quantitative Turn in Journalism?* TOW CENTER FOR DIGITAL JOURNALISM (Oct. 30, 2013), <http://towcenter.org/blog/a-quantitative-turn-in-journalism/>; Mark Coddington, *Clarifying Journalism’s Quantitative Turn: A Typology for Evaluating Data Journalism, Computational Journalism, and Computer-Assisted Reporting*, 3 DIGITAL JOURNALISM 331, 348 (2015).

⁴⁸ Washington PostPR, *The Washington Post Establishes A Computational Political Journalism R&D Lab to Augment Its Campaign 2020 Coverage*, WASH. POST (July 24, 2019), <https://www.washingtonpost.com/pr/2019/07/24/washington-post-establishes-computational-political-journalism-rd-lab-augment-its-campaign-coverage/>.

⁴⁹ Wall Street Journal, *The Wall Street Journal is Hiring Engineers*, MEDIUM (Aug. 6, 2017), <https://medium.com/@WSJ/the-wall-street-journal-is-hiring-engineers-9c18e17e163a> (example of one of several hiring posts in recent years).

⁵⁰ Benjamin Mullin, *Bloomberg EIC: Automation is ‘Crucial to the Future of Journalism,’* POYNTER (Apr. 27, 2016), <https://www.poynter.org/tech-tools/2016/bloomberg-eic-automation-is-crucial-to-the-future-of-journalism/>.

⁵¹ Katherine Fink & Christopher W. Anderson, *Data Journalism in the United States: Beyond the “Usual Suspects,”* 16 JOURNALISM STUDIES 467, 481 (2015).

⁵² NICHOLAS DIAKOPOULOS, *AUTOMATING THE NEWS: HOW ALGORITHMS ARE REWRITING THE MEDIA* (2019).

⁵³ Meredith Broussard, *Artificial Intelligence for Investigative Reporting: Using an Expert System to Enhance Journalists’ Ability to Discover Original Public Affairs Stories*, 3 DIGITAL JOURNALISM 814, 831 (2015).

that no one today has time or money to mine. With a suite of reporting tools, a journalist will be able to scan, transcribe, analyze, and visualize the patterns in these documents.”⁵⁴

However, even as data collection becomes ubiquitous throughout society, both journalists and scholars have mounted concerns over attempts by governments to exercise stricter control over and limit access to data. Although public records requests were often met with governmental restraint and suppression tactics long before the rise of computational journalism,⁵⁵ data access issues pose unique obstacles for the implementation of computational methods and the realization of its benefits. For example, a necessary component to apply computational methods is access to structured data. Structured data consists of clearly defined data types whose pattern makes them easily searchable. Yet, a common refrain from journalists is how responses to record requests often include boxes of printed emails or PDF screenshots of Excel files.⁵⁶ Although there exist journalism-specific tools for analyzing unstructured documents such as Tabula, sifting through and scanning paper documents or other unstructured documents takes valuable resources.⁵⁷ Furthermore, as stated above, requests for data stored in governmental databases are often met with several arguments as to why access should be denied or require exorbitant fees. Such responses may require abundant resources in terms of money, lawyers, and time, thereby constraining computational journalism to only those news organizations that can afford it.⁵⁸

Previous empirical research

This study is premised, in part, on the relative lack of empirical research examining the public accessibility of structured data. However, scholars have published related work. The closest in subject and method is Robert Goodspeed, who examined how Massachusetts municipalities respond to public records requests for geographic information systems (“GIS”) data.⁵⁹ The study entailed mailing records requests to 351 cities and towns in the state and then analyzing the outcomes, such as how many municipalities responded (278, or 78.6%), response times, fees charged, whether GIS records were provided in a digital format, arguments asserted for nondisclosure, and attempts by the municipalities to restrict use of the data.⁶⁰ Most of the municipalities responded to the request, with an average response time of less than seven days, and 20.3% did not impose a fee. Some municipalities also denied the request, refused to produce the data in electronic format, or imposed use restrictions.⁶¹ The study concluded: “Open government will require not only moving beyond a narrowly conceived ‘right to view’ records but also an ongoing discussion of a range of practical, ethical, and legal considerations such as file formats and distribution technology, balancing openness with privacy and public safety considerations, and appropriate data policies and licenses.”⁶²

Also relevant to this project is legal research synthesizing judicial decisions about public access to government data. Much of the literature in this realm is from the early 1990s and entails

⁵⁴ Sarah Cohen, James T. Hamilton, & Fred Turner, *Computational Journalism*, 54 COMM. OF THE ACM 66 (2011).

⁵⁵ See e.g., Martin E. Halstuk & Bill F. Chamberlin, *The Freedom of Information Act 1966–2006: A Retrospective on the Rise of Privacy Protection Over the Public Interest in Knowing What the Government’s Up To*, 11 COMM’N. L. & POL’Y (2006).

⁵⁶ Jonathan Stray, *Making Artificial Intelligence Work for Investigative Journalism*, 7 DIGITAL JOURNALISM 1076, 1097 (2019).

⁵⁷ *Id.*

⁵⁸ *Id.*

⁵⁹ Robert Goodspeed, *From Public Records to Open Government: Access to Massachusetts Municipal Geographic Data*, 23 URISA JOURNAL 21 (2011).

⁶⁰ *Id.*

⁶¹ *Id.*

⁶² *Id.* at 30-31.

either legal analysis of what the law is or prescriptive opinions about what the law should be. The studies have historical value, but their conclusions are not current given legislative changes and new case law. Among the more comprehensive analyses is from Bunker, Spilchal, Chamberlin, & Perry (1992), who reviewed court decisions at the state and federal levels governing access to databases and other forms of electronic records.⁶³ They found that court rulings on the issue of access to computer records was mixed, with some courts disagreeing about whether agencies had to make data available in a format desired by the requester.⁶⁴ The article also listed 13 recommendations for reforming and updating access laws, such as requiring agencies to publicly disclose what kinds of data are maintained in government computers.⁶⁵ “Access to computerized information presents different problems than those associated with access to paper files,” the authors concluded.⁶⁶ “If would-be requesters do not begin to demand the same level of access to computer files they now have to paper files, hard-won privileges of access may vanish.”⁶⁷

Another dimension of related research is empirical in nature but focuses on the extent to which the federal government has complied with the Electronic Freedom of Information (“EFOIA”) Act of 1996, which, among other things, clarified that records maintained electronically were subject to FOIA, and required agencies to post policy documents and frequently requested records online.⁶⁸ Notable studies in this category are from Gordon-Murnane (1999)⁶⁹ and Oltmann, Rosenbaum, & Hara (2007),⁷⁰ who performed content analyses of federal agency websites. Both studies found a lack of legal compliance, including failure by agencies to post required records online and include a link on their main website to facilitate FOIA requests.⁷¹

A closely related form of research to our project entails analyzing FOIA logs or third-party record requests to study how public records laws are used. Cuillier (2019) analyzed data on more than 7,000 records requests filed with state agencies through the website MuckRock.com and then tested for correlations between the extent of disclosure and various legal, political, demographic, and cultural attributes.⁷² Fink (2017) examined responses to federal FOIA requests, including her own requests, to assess whether federal agencies would release algorithms.⁷³ Fink (2018) examined public records requests filed with 21 state environmental agencies, or logs of such requests, and found that journalists accounted for only a sliver of overall requests.⁷⁴ Kwoka (2016) has come to a similar conclusion at the federal level after examining FOIA logs from various agencies, concluding that commercial and first-person requesters often are the most frequent FOIA filers, although it depends

⁶³ Matthew D. Bunker, Sigman L. Spilchal, Bill Chamberlin & Linda M. Perry, *Access to Government-Held Information in the Computer Age: Applying Legal Doctrine to Emerging Technology*, 20 FLA. ST. U. L. REV. 543 (1993).

⁶⁴ *Id.*

⁶⁵ *Id.*

⁶⁶ *Id.* at 24.

⁶⁷ *Id.*

⁶⁸ Martin E. Halstuk, *Speed Bumps On The Information Superhighway: A Study Of Federal Agency Compliance With The Electronic Freedom of Information Act Of 1996*, 5 COMM. L. & POL’Y 423 (2000).

⁶⁹ Laura Gordon-Murnane, *The Electronic Freedom of Information Act: The Promise Remains Unfulfilled*, 7 SEARCHER 34 (1999).

⁷⁰ Shannon M. Oltmann, Howard Rosenbaum, & Noriko Hara, *Digital Access to Government Information: To What Extent Are Agencies in Compliance with EFOIA?*, 43 PROCEEDINGS OF THE AM. SOC. FOR INFO. SCI. & TECH. 1 (2006).

⁷¹ *Id.*

⁷² David Cuillier, *Bigger Stick, Better Compliance? Testing Strength of Public Record Statutes on Agency Transparency in the United States* 8 (June 26, 2019) (paper presented to the Global Conference on Transparency Research, Rio de Janeiro, Brazil), <https://drive.google.com/file/d/1R9PBjrjTreetcD-epJEq-o8ZSWZPKRjv/view>.

⁷³ Katherine Fink, *Opening the Government’s Black Boxes: Freedom of Information and Algorithmic Accountability*, 21 INFO, COMM. & SOC’Y 1453 (2017).

⁷⁴ Katherine Fink, *State FOI Laws: More Journalist-Friendly, or Less?*, in TROUBLING TRANSPARENCY: THE HISTORY AND FUTURE OF FREEDOM OF INFORMATION (David E. Pozen & Michael Schudson eds., 2018).

on the agency.⁷⁵ Using natural language processing, Berliner, Bagozzi, & Palmer-Rubin (2018) analyzed the text of more than one million records requests filed with Mexican federal agencies to investigate what kinds of information requesters were seeking.⁷⁶ The study found that 14 of the 20 primary topics of records requests implicated government accountability. Roberts (2002) examined data covering 2,120 records requests filed with a Canadian federal agency and found that requests deemed politically sensitive or that were from the media or political parties took longer to process.⁷⁷

Finally, the methodological model for our study is the access audit, which involves filing public records requests with government agencies to evaluate how the agencies respond.⁷⁸ Such studies may take the form of experiments when they have a control group and treatment group, but not all audits are designed this way. Still, audits give researchers greater control over records requests—such as how they are worded, the government agencies involved, and the extent of interaction with record custodians—than analyses of FOIA logs or third-party record requests, although there are benefits and drawbacks to both. While the benefit of access audits is control over the requests, a significant drawback may be that they may not fully reflect how government officials provide information to average people in everyday settings, such as the local resident who calls the village clerk or sits in on a town board meeting. Indeed, there is evidence that record custodians do not always follow the strict letter of the law when deciding whether to release information to the public.⁷⁹ On the other hand, the chief benefit of analyzing FOIA logs is that they reflect “real” records requests and responses.

Audits have been used domestically and around the world. In the United States, Cuillier (2010) sought police use-of-force reports and school superintendent employment contracts to test whether the tone of records requests—friendly, neutral, or threatening—affected how an agency responded.⁸⁰ The study found that a threatening letter was more effective at obtaining records, and also had faster response times and lower fees, than the friendly and neutral tones.⁸¹ Fink (2019) submitted public records requests for dog license records to municipalities in New York to evaluate the quantity and nature of interactions with record custodians.⁸² The study showed that multiple interactions were often required to obtain information after filing records requests, which may pose a challenge for community journalists who seek to balance obtaining information with maintaining

⁷⁵ Margaret B. Kwoka, *FOIA, Inc.*, 65 DUKE L.J. 1361 (2016); Margaret B. Kwoka, *The Other FOIA Requesters*, in TROUBLING TRANSPARENCY: THE HISTORY AND FUTURE OF FREEDOM OF INFORMATION (David E. Pozen & Michael Schudson eds., 2018); Margaret B. Kwoka, *First-Person FOIA*, 127 YALE L.J. 2204 (2018).

⁷⁶ Daniel Berliner, Benjamin E. Bagozzi, & Brian Palmer-Rubin, *What Information Do Citizens Want? Evidence from One Million Information Requests in Mexico*, 109 WORLD DEVELOPMENT 222 (2018).

⁷⁷ Alasdair Roberts, *Administrative Discretion and the Access to Information Act: An “Internal Law” on Open Government*, 45 CAN. PUB. ADMIN. 175 (2002).

⁷⁸ Access audits are also conducted frequently by good government groups and organizations that advocate for the legal right of access to government information, but we do not review that work here. One of the earliest such audits was conducted by Ralph Nader using FOIA shortly after it became law. See Ralph Nader, *Freedom from Information: The Act and the Agencies*, 5 HARV. C.R.-C.L. L. REV. 1 (1970); David E. McCraw, *The “Freedom From Information” Act: A Look Back at Nader, FOIA, and What Went Wrong*, 126 YALE L.J. F. 232 (2016).

⁷⁹ Brett G. Johnson, *Public Records Officers’ Perspectives on Transparency and Journalism*, 2 J. CIVIC INFO., 1 (2020); Michele Bush Kimball, *Public Records Professionals’ Perceptions of Nuisance Requests for Access*, 5 U. BALT. J. MEDIA L. & ETHICS 46 (2016); Michele Bush Kimball, *Shining the Light from the Inside: Access Professionals’ Perceptions of Government Transparency*, 17 COMM. L. & POL’Y 299 (2012); Michele Bush Kimball, *Law Enforcement Records Custodians’ Decision-Making Behaviors in Response to Florida’s Public Records Laws*, 8 COMM. L. & POL’Y 313 (2003).

⁸⁰ Cuillier, *supra* note 12.

⁸¹ *Id.*

⁸² Fink, *supra* note 12.

relationships with officials.⁸³ Wagner (2020) submitted 1,002 requests to agencies in nine states for records of sheriff reports, pothole complaints, county collective bargaining agreements, and in some cases, budget information.⁸⁴ The study identified geographic location, race, and political attributes as significant variables in the outcomes of records requests.⁸⁵ ben-Aaron, Denny, Desmarais, & Wallach (2017) used a randomized field experiment to test whether record custodians in North Carolina counties were more likely to fulfill a public records request if they knew that their peers in other counties had also fulfilled the request. The study found an effect: Notifying record custodians that a records request was fulfilled elsewhere increased the production of information and decreased wait times. Audits have also been conducted at the federal level using FOIA.⁸⁶

Internationally, Worthy, John, & Vannoni (2017) found that invoking a public records law was more effective at getting information out of English parish councils than informal requests. Grimmelikhuijsen, John, Meijer, & Worthy (2018) replicated Worthy, et al. (2017) by conducting a similar experiment in the Netherlands, finding that invoking the law led to greater disclosure than informal requests. Lagunes & Pocasangre (2019) examined the performance of Mexico's Freedom of Information Act over eight years by submitting 307 records requests to 197 federal government entities. They in part sought to test whether there would be a difference in responses based on whether the government perceived the requester as an average citizen or someone with "political and economic clout," but found no such effect. The study confirmed the results of a similar study from years earlier also conducted by Lagunes (2006).⁸⁷

Research questions

Given the importance of public access to structured data held by the government, this study uses an audit to answer the following questions:

- RQ1:* To what extent do public universities release structured data in response to public records requests?
- RQ2:* What reasons do public universities give for not disclosing structured data?
- RQ3:* To what extent do public universities require prepayment for processing records requests that seek structured data?
- RQ4:* What file formats are commonly released by public universities in response to public records requests for structured data?
- RQ5:* How long does it take for public universities to respond to public records requests for structured data?

⁸³ *Id.*

⁸⁴ Wagner, *supra* note 12.

⁸⁵ *Id.*

⁸⁶ David E. Lewis & Abby K. Wood, *The Paradox of Agency Responsiveness: A Federal FOIA Experiment* (April 2012) (paper presented at the Midwest Political Science Association annual meeting), https://www.vanderbilt.edu/csdi/research/CSDI_WP_06-2012.pdf/.

⁸⁷ Paul Lagunes, *Irregular Transparency? An Experiment Involving Mexico's Freedom of Information Law* (SSRN Scholarly Paper No. ID 1398025, 2009).

Method

To better understand the potential issues with accessing databases and structured data, we filed public records requests with the largest public university⁸⁸ of each state between June and July 2020, except in six states with requester residency requirements,⁸⁹ leaving 44 states for testing. This resulted in 132 requests. We selected public universities because of the likelihood that they maintain common types of databases and because each state has a public university system. We chose the largest public university by enrollment in each state to measure the performance of substantially similar agencies. Private universities are generally not subject to public records laws, although their police departments are in some states.⁹⁰

Each request sought three records: (1) data on payments to vendors; (2) data on payroll payments to employees; and (3) an email log of the university's most senior administrator, typically the university president or chancellor.⁹¹ These data were selected because they implicate a range of access issues. We assumed that vendor payments would be the easiest data to obtain; vendor payments are likely to be maintained as structured data in accounting software, and it is widely accepted in the United States that the public generally has a right to know how the government spends public money. We further assumed that payroll data would be reasonably accessible because such information also is likely maintained in a database, but that there could be some challenges in obtaining the information as discrete payroll payments instead of aggregate amounts or annual salaries. Moreover, we assumed that universities might attempt to withhold identifying information for student workers pursuant to the Family Educational Rights and Privacy Act ("FERPA").⁹² Finally, we assumed that the request for email logs would likely be the most challenging of the requests. While full emails are routinely produced in response to public records requests, finding and/or generating email logs may require advanced technical knowledge beyond the expertise of record custodians, and indeed may require consultation with email service providers or information technology staff. Moreover, we assumed that email logs would be the most likely of the three requests to be redacted, which can be complicated when working with structured data.

To make the records requests, we obtained a listing of the largest public universities of each state by enrollment and then filed requests using both a template and/or online portals when appropriate.⁹³ Preliminary research was done on each university to assess how requests should be filed, principally by using a search engine with the university's name followed by the words "public records request." Requests were filed either by emailing designated record custodians as listed on university websites; emailing senior communication, legal, and/or administrative officials; or by using online portals that some universities operate to handle public records requests. The request included a caveat that if processing exceeded a defined amount, written authorization would be needed before proceeding. When a university sought to impose a fee before processing, the

⁸⁸ The largest public university in Pennsylvania is The Pennsylvania State University, however, it is largely exempt from the state's Right-To-Know Law. As such, a request was filed with the largest public university that is subject to the Right-To-Know Law, West Chester University of Pennsylvania.

⁸⁹ States with requester residency requirements are Alabama, Arkansas, Delaware, Georgia, Tennessee, and Virginia. These states allow agencies to deny a request if the requester is not a resident of the state. The Supreme Court of the United States has upheld the constitutionality of such provisions. *See* *McBurney v. Young*, 569 U.S. 221 (2013).

⁹⁰ Josh Moore, *Out from the Curtains of Secrecy: Private University Police and State Open Records Laws*, 2 J. CIVIC INFO 1 (2020).

⁹¹ *See* Appendix A for an example of the public records request.

⁹² 20 U.S.C. § 1232g.

⁹³ The listing of the of the largest public universities of each state was obtained from the Department of Education National Center for Education Statistics.

university's response was coded as requiring a fee before processing; the fee was not paid and the records were no longer sought.

We took steps to mitigate against any undue influence on how universities responded. These steps included: (1) not revealing that the requests were part of a research project evaluating universities' responses; (2) responding to any questions or inquiries from universities within 24 hours; (3) not affirmatively contacting universities about the status of requests, except for a one-time reminder inquiry for universities that had not provided a substantive response to any of the three requests after 66 business days; and (4) agreeing to use existing online datasets to fulfill requests when referred to them by a university and when the data were substantially similar to the requested record.

When a university provided a substantive response,⁹⁴ the following variables were coded:

- **Nature of response:** Whether the university produced responsive data, produced or offered different data than what we requested, asserted there were no records, required prepayment before processing, required in-person inspection, or denied the request.
- **Response time:** The number of business days (i.e., omitting weekends and holidays) from the day after a request was filed to the day a substantive response was received.
- **Format:** The format that data were released: Excel, CSV, PDF, or web page.
- **New record:** Whether the university expressly asserted that it was not obligated to create a new record in response to a public records request.
- **Fee estimate:** The amount of money a university estimated it would cost to process the request.

Results⁹⁵

Our first research question asked: To what extent do public universities release structured data in response to public records requests? Analysis of the study's results show that the universities in our sample produced responsive structured data without a fee in slightly more than a quarter of requests ($N=37$, 28%). That is, universities released exactly what we requested in a structured format and at no charge about a quarter of the time.

Overall, we received substantive responses to 110 (83%) of the 132 requests filed. As Table 1 shows (next page), we make two types of distinctions among records that were released. First, we separate responses that produced structured data from responses that produced records in a non-structured format. Among all records released, 43 (74%) were in a structured data format and 15 (26%) were in a non-structured format. Second, we distinguish between records that were responsive (i.e., containing the information we requested) and similar records, which did not meet precise request specifications but were otherwise similar to the information we sought. Of records released, 42 (72%) resulted in access to responsive records, while another 16 requests (28%) resulted in the production or offering of similar records.

⁹⁴ By "substantive response" we mean a response in which a university produced records, explained why it could not or would not produce records, or required prepayment before processing.

⁹⁵ The dataset we created for this project is available online at <https://hdl.handle.net/11299/219041>.

Table 1. Overall Response

	University Response								
	Produced Structured Data		Produced Non-Structured Record		No Responsive Records	Required Prepayment Before Processing	Required In-Person Inspection	Denied	No Substantive Response
	Responsive Records	Similar Records	Responsive Records	Similar Records					
Vendor Payments (N=44)	20	1	2	5	5	6	0	0	5
Payroll Payments (N=44)	11	5	1	5	6	7	1	1	7
Email Logs (N=44)	6	0	2	0	15	10	0	1	10
Total (N=132)	37	6	5	10	26	23	1	2	22

Vendor payments were the most accessible. Twenty-two universities produced records responsive to our requests and six produced or offered to produce records that were similar but not responsive in six requests.⁹⁶ An example of a record not responsive is a spreadsheet of vendor payments we received from North Dakota State University, which listed payments outside the time period of our request. Twenty-one universities produced vendor payment records in a structured data format, while seven universities released non-structured records. Five universities asserted there were no responsive records and six universities required prepayment before processing. We did not receive substantive responses from five universities for vendor payment records.

Payroll payments were the second-most accessible, with 12 universities disclosing responsive records and 10 universities providing similar records, namely aggregate salary information instead of records of every payroll period covered by the request. Sixteen universities produced records in a structured data format, while six universities released records in a non-structured format. Six universities asserted they had no responsive records and seven universities required prepayment. Louisiana State University required inspection of payroll records on campus,⁹⁷ even though Louisiana's Public Records Act expressly permits copies of public records,⁹⁸ and

⁹⁶ The University of Nebraska asserted it had no records responsive to any of our requests, but provided a link to a state transparency website that the records custodian said shows “[p]ayments made by the Treasurer of the State of Nebraska on behalf of the University of Nebraska.” However, the website does not allow users to query data deriving from the university or include payroll payments. Because none of the data could be associated with the university, this response was coded as No Responsive Records and not Produced/Offered Different Data.

⁹⁷ Email from Tetyana Hoover, Louisiana State University, to Sarah K. Wiley (July 10, 2020) (on file with authors).

⁹⁸ LA. STAT. ANN. § 44:31.

Rutgers asserted that our request was overly broad.⁹⁹ Seven universities did not provide a substantive response.

By comparison, administrator email logs were the least accessible. Only eight universities produced email logs, six of which were in a structured format. Meanwhile, 15 universities asserted they did not maintain email logs,¹⁰⁰ 10 required prepayment before processing the request for email logs, and the University of Kentucky said the request was overly broad.¹⁰¹ Ten universities did not provide a substantive response.

Our second research question asked what reasons public universities gave for not disclosing structured data. In response to 26 of the requests, universities said they did not have any responsive records; universities in 21 requests also expressly asserted that they were not required to create new records. These were the most common reasons given for nondisclosure. Some universities made the claims for all three requests, while other universities made the claims in response to just one of the requests. As we discuss later, this position is particularly significant in the context of structured data, which may be maintained in ways that authorities do not always recognize as potentially responsive to records requests.

Language of the response letters is illustrative. The University of Missouri-Columbia's record custodian wrote: "As you may be aware, the Missouri Sunshine Law does not require that a public governmental body create a document where such a document does not already exist; rather, it provides a mechanism for the public to obtain copies of existing records. The University does not have existing records of the three types your request describes and we decline to create such reports."¹⁰² An official at the University of California-Los Angeles wrote that while the university had conducted "a reasonable and good faith search, we have been unable to locate any existing responsive records. Moreover, while the [California Public Records Act] requires public agencies to provide access to certain existing records that are not exempt from disclosure, it does not require them to create new records to satisfy a request."¹⁰³ The response then cited two state appellate court decisions to justify its response: *Sander v. State Bar of California*, 26 Cal.App.5th 651 (2018) and *Fredericks v. Superior Court* (2015) 233 Cal. App. 4th 209, 227.¹⁰⁴ The University of Nebraska's records custodian wrote: "the University does not have an existing compilation report responsive to your request and would have to create a new record in order to provide you with this information. Pursuant to Neb. Rev. Stat. § 84-712(3)(e),¹⁰⁵ the University is not required to create a new record or produce the records in a different form or format modified from that of the original public record."¹⁰⁶

⁹⁹ Email from Jewell Battle, Rutgers University, to Jonathan Anderson (Sept. 1, 2020) (on file with authors).

¹⁰⁰ The University of New Mexico offered to produce copies of emails instead of a log. We declined the offer.

¹⁰¹ Email from Brenda McFaddin, University of Kentucky, to Sarah K. Wiley (Aug. 21, 2020) (on file with authors). McFaddin wrote, in relevant part: "Although 'K.R.S. § 61.872(2) only requires that one seeking to inspect public records may be required to submit a written application 'describing the records to be inspected,' the written request must still be 'adequate for a reasonable person to ascertain the nature and scope of [the] open records request.' *Commonwealth v. Chestnut*, 250 S.W.3d 655, 661 (KY. 2008). Agencies are not required to speculate as to what the requester wants. Nor are agencies required to assume the breadth of a request. You must specifically name each University employee and/or provide their email address for the time frame you previously provided. Additionally, please provide specific search terms so that our IT Department may conduct the appropriate searches for this request. Thank you."

¹⁰² Email from Paula Barrett, University of Missouri-Columbia, to Sarah K. Wiley (Oct. 27, 2020) (on file with authors).

¹⁰³ Email from Rachel Pounds, University of California-Los Angeles, to Sarah K. Wiley (Aug. 10, 2020) (on file with authors).

¹⁰⁴ *Id.*

¹⁰⁵ NEB. REV. STAT. § 84-712(3)(e) states: "This section shall not be construed to require a public body or custodian of a public record to produce or generate any public record in a new or different form or format modified from that of the original public record."

¹⁰⁶ Email from Erin E. Busch, University of Nebraska, to Jonathan Anderson (July 6, 2020) (on file with authors).

Most of the records that were released did not contain redactions. Of those that did, common justifications for nondisclosure were student privacy laws, including the Family Educational Rights and Privacy Act. The University of Washington asserted that it was required to redact student names in payroll data pursuant to FERPA.¹⁰⁷ Louisiana State University wrote that “FERPA requires nondisclosure of student information” in the email logs, although the university did not provide a substantive response for such records.¹⁰⁸ Texas A&M University redacted records pursuant to student privacy protections as well as an exemption covering competitive information for “[s]alary [i]ncome.”¹⁰⁹

Finally, 22 universities did not provide a substantive response. Four universities failed to provide a substantive response to all three requests: University of Alaska, Anchorage; Arizona State University, Tempe; University of Maryland, University College; and University of Oklahoma, Norman. The University of Kentucky, North Dakota State University, and Oregon State University produced records in response to one request but did not respond to the other two. Four universities responded to two requests but not a third: South Dakota State University; Louisiana State University; University of Washington, Seattle¹¹⁰; and University of Wisconsin, Madison.

Our third research question asked: To what extent do public universities require prepayment for processing records requests that seek structured data? Table 2 (next page) shows all fee estimates received from universities, which varied widely. While data in 58 requests were produced without charge, 14 universities estimated fees associated with 25 requests.¹¹¹ The fee estimates ranged from slightly more than \$50 to upwards of \$1,000. The public records officer at Iowa State University, which estimated the largest fee as between “several hundred dollars to possibly over \$1,000,” wrote that it was difficult to estimate the amount of time it might take to produce records.¹¹² “Your records request is complicated and time-consuming for us and will require pre-payment,” the public records officer wrote.¹¹³ She described the requests for vendor and payroll data as “the onerous parts,” but that the request for the email log “would take us only about an hour.”¹¹⁴ Such a response was not the norm; other universities released the vendor and payroll data easily but said the email logs were difficult. Indeed, fees were most prevalent for the email log, which were triggered in 11 requests, although that figure does not include any of the 10 universities that did not respond to the email request. Fees were triggered in seven requests for both the vendor and payroll data.

¹⁰⁷ Letter from Jade McNallan, University of Washington, to Jonathan Anderson (Aug. 31, 2020) (on file with authors).

¹⁰⁸ Email from Tetyana Hoover, Louisiana State University, to Sarah K. Wiley (July 10, 2020) (on file with authors).

¹⁰⁹ Email from Tricia Bledsoe, Texas A&M University, to Jonathan Anderson (Aug. 12, 2020) (“[R]edactions have been made in accordance with Section 552.104 Competitive Information (Santucci Salary Income) and Section 552.114 of the Texas Public Information Act.”) (on file with authors).

¹¹⁰ The University of Washington, Seattle may have intended to respond to all three requests, but one of its mailings to us contained a CD with records for a different requester.

¹¹¹ The count of 25 requests that triggered fee estimates in Table 2 is greater than the 23 requests in which universities required prepayment before processing, as shown in Table 1. This discrepancy is because fee analysis is decoupled from response analysis. The goal of the fee analysis is to assess the quantity and nature of fees that universities asserted, while the response analysis in Table 1 reflects a basic typology of request outcomes. Here, responses from two universities caused the incongruity. Iowa State University’s response referred to online datasets that were similar, but not exactly responsive to, our requests for vendor and payroll data. Pursuant to our coding scheme for Table 1, we coded Iowa State’s response to vendor and payroll requests as producing/offering different data. However, for purposes of assessing fee estimates, we considered Iowa State’s full fee estimate, which applied to the vendor and payroll data requests and not inexact data posted online. Moreover, the University of Montana said there were no responsive records for the email log, but the university offered to manually create a log for an unspecified fee. In short, this explains why Table 2 identifies 25 requests associated with fee estimates and Table 1 identifies 23 requests as requiring prepayment before processing.

¹¹² Email from Sheryl Rippeke, Iowa State University, to Sarah K. Wiley (July 21, 2020) (on file with authors).

¹¹³ *Id.*

¹¹⁴ *Id.*

Table 2. Fee Estimates

University	Fee Estimate	Records Covered V=Vendor data P=Payroll data E=Email log
Iowa State University	"[S]everal hundred dollars to possibly over \$1,000." ¹¹⁵	V, P
Utah Valley University	\$961.51	E
University of Wisconsin-Madison	\$496.24-\$951.24	V, P
University of Massachusetts, Amherst	\$800	V, P, E
University of Kansas	\$622	V, P, E
Michigan State University	\$280	V
University of Colorado-Boulder	\$235.06	E
Rutgers University, New Brunswick ¹¹⁶	\$179.10	E
University of Rhode Island ¹¹⁷	\$165	V, P, E
University of Wyoming ¹¹⁸	\$90	V, P, E
University of South Carolina	\$75	E
University of Central Florida	\$53.66	P, E
University of Montana	Not specified	E
University of Minnesota-Twin Cities	Not specified	E

Our fourth question asked: What file formats are commonly released by public universities in response to public records requests for structured data? Table 3 (next page) shows that data released in response to 43 requests were in a structured format. Data released in 37 requests were Excel files, while there were six requests that resulted in comma-separated value files. The remaining data were produced in non-structured formats: universities released PDF files in response to seven

¹¹⁵ Email from Sheryl Rippke, Iowa State University, to Sarah K. Wiley (July 21, 2020) (on file with authors).

¹¹⁶ The gross fee estimate from Rutgers University was \$359.10 for 7.98 hours of time to review and potentially redact the records. However, the university said it would waive the first four hours, thereby reducing the fee estimate to \$179.10.

¹¹⁷ The University of Rhode Island's total fee estimate was \$180, but pursuant to 38 R.I. GEN. LAWS ANN. § 38-2-4 (West), the first hour is waived. Thus, the total fee estimate was \$165.

¹¹⁸ The University of Wyoming's total fee estimate was \$270, but pursuant to university policy, fees are charged only if they exceed \$180, and only the amount exceeding \$180.

requests and referred to web pages without a data download function in response to eight requests. Where an online dataset could be downloaded as a structured data file (e.g., CSV), the format was coded as that file type. If the online dataset could not be downloaded, the data were coded as a web page. The requests sought responsive data as Excel spreadsheets, comma-delimited text files, or if neither format was possible, the existing format of the data. Some of the PDF files clearly were created from underlying Excel files, including screenshots of Excel files. Four universities provided responsive records for all three requests in a structured format: College of Southern Nevada, University of Illinois at Urbana-Champaign, West Chester University of Pennsylvania, and Texas A&M, College Station.

Table 3. Format of Data Provided

	Format			
	CSV	Excel	PDF	Web Page
Vendor Data	2	19	3	4
Payroll Data	3	13	2	4
Email Log	1	5	2	0
Total	6	37	7	8

Our last question asked: How long does it take for public universities to respond to public records requests for structured data? The median response time for all requests that received substantive responses was 16 days. This was also the median for vendor and payroll requests, while the median response time for email logs was 19.5 days. As Table 4 (next page) shows, universities issued final responses within 10 business days in 36% of processed requests, with 70% of final responses issued within 30 business days. The median response time to provide structured data responsive to our requests was 23 business day, while the median response time for non-structured records was 11 business days. Sixteen of the universities that responded fairly promptly are subject to public records laws that require records custodians to take some sort of action toward processing a public records request within 10 or fewer business days of receiving the request.¹¹⁹

For universities that had not provided a substantive response to any of the three requests after 66 business days ($N=14$), we emailed a reminder inquiry asking about the status of the request. As of March 9, 2021, the oldest unfilled request had been pending for 169 business days. After the reminder inquiry, some universities produced responsive records, while other universities had yet to do so as of the time of publication. Universities that received reminder inquiries are noted in Appendix B.

¹¹⁹ HAW. CODE R. 2-71-13 (West); IDAHO CODE ANN. § 74-103 (West); 5 ILL. COMP. STAT. ANN. 140/3; IND. CODE ANN. § 5-14-3-9 (West); KAN. STAT. ANN. § 45-218 (West); LA. STAT. ANN. § 44:32; NEV. REV. STAT. ANN. § 239.0107 (West); N.M. STAT. ANN. § 14-2-8 (West); OR. REV. STAT. ANN. § 192.324 (West); 38 R.I. GEN. LAWS ANN. § 38-2-3 (West); S.C. CODE ANN. § 30-4-30; S.D. CODIFIED LAWS § 1-27-37; UTAH CODE ANN. § 63G-2-204 (West); VT. STAT. ANN. TIT. 1, § 318 (West); W. VA. CODE ANN. § 29B-1-3 (West); WYO. STAT. ANN. § 16-4-202 (West).

Table 4. Response Time

	Response Time (Days)					Pending Requests
	1-10	11-30	31-45	46-90	≥91	
Vendor Data	15	13	2	7	2	5
Payroll Data	14	11	3	8	1	7
Email Log	11	13	2	7	1	10
Total	40	37	7	22	4	22

Discussion

The research reported in this article reveals significant obstacles to obtaining structured data under state public records laws. The universities in our sample produced responsive structured data without a fee in slightly more than a quarter of requests ($N=37$, 28%), meaning the vast majority of our requests ($N=95$, 72%) failed to yield the information we sought in a structured data format and for free. This is particularly concerning given that at least two of the three requests were for fairly routine data, vendor payments and payroll payments. Moreover, the median response time for responsive structured data was 23 business days, but ranged widely, between four business days and up to 111 business days, with some universities neglecting to respond until only after we inquired about the status of our requests. Universities in 22 requests failed to provide any substantive response at all. Of the 23 requests that triggered specific fees, 17 had fee estimates of more than \$100 and several were near \$1,000.¹²⁰

Another troubling finding was that universities repeatedly asserted there were no records responsive to our request or that the request would require the creation of a new record. Universities asserted there were no responsive records to 21 of our requests, but such a response is suspect given that any large enterprise, especially as a governmental entity, is likely to maintain the requested data. Moreover, universities using similar technologies responded differently. For example, two universities that released email logs—University of Illinois at Urbana-Champaign and Boise State University—use popular email systems. Illinois uses Microsoft¹²¹ and Boise State uses Gmail.¹²² Other universities in our sample also use these same types of email systems, yet they asserted that email logs did not exist.¹²³ The repeated responses from universities that no records existed demonstrates the need for greater public disclosure about what databases government agencies use and the capabilities of those databases for retaining and producing data.

Some universities that denied our requests based on the assertion that it would require them to produce a new record are located within states where either case law or statutes allow for such a conclusion—or leaves room for poor interpretation. For example, there is no explicit statutory

¹²⁰ Universities sought to impose fees in 25 requests but did not specify the fee amount in two requests.

¹²¹ Email, Information Technology Services, University of Illinois, <https://techservices.illinois.edu/services/email>.

¹²² Gmail at Boise State, Boise State University, <https://www.boisestate.edu/oit-gsuite/gmail-at-boise-state-university/>.

¹²³ *E.g.*, The University of New Mexico, the University of Vermont, and Indiana University all use Microsoft email systems for faculty and staff, and the University of Hawaii uses Gmail.

authority for a requester to obtain a customized search of an electronic database in Vermont.¹²⁴ However, the Vermont Supreme Court has recognized that “[a]s a practical matter, the steps required to reasonably compile requested public records” may include “centralized electronic searches of agency records in an email system, document management application, or database[s] within specified parameters” and that such searches “may be the primary or even exclusive means of compiling responsive public records.”¹²⁵ As stated above, the University of Vermont’s response denying our request stated “Vermont’s Public Records Act does not oblige the University to generate custom reports or create new records.”¹²⁶

Similarly, although there are no state Supreme Court decisions on point in West Virginia, the Kanawha Circuit Court found that while a public body is not required to create a new public record, “neither the computer program used to search an electronic database, nor the results of a search of an electronic database constitute the creation of a new public record.”¹²⁷ West Virginia University’s response denying our request stated, “There is no obligation under the State FOIA to create any particular record, but only to provide access to a public record already created.”¹²⁸ Such responses from West Virginia University and the University of Vermont highlight how case law within the respective states leaves room for poor interpretation.

Furthermore, although New Mexico and Hawaii both have statutes addressing electronic databases, both statutes leave room to deny requests if it requires a “public body to create a public record”¹²⁹ or, if in Hawaii, the information is not “readily retrievable from the database in question.”¹³⁰ New Mexico’s Inspection of Public Records Act also specifically states that a customized search of a database is discretionary.¹³¹ University of Hawaii’s response denying our request stated that records could not be provided because the request required the university to “create a summary or compilation from records, but requested information is not readily available.”¹³² Meanwhile, the University of New Mexico’s record custodian stated that agencies do not have to generate new records if doing so requires more than a few keystrokes.¹³³

As our responses highlight, absent explicit guidance—either within case law or by statute—public institutions deny access to information contained within electronic databases because doing so would require them to “create a new record.” To be clear: this is the same information that if recorded on paper and located within file cabinets public institutions would be required to disclose. Such denials are troubling, especially considering the increased collection and storage of data within electronic databases.

In December 2020, the Ninth Circuit was the latest federal court to address this issue, ruling that querying a database did not in fact constitute the creation of a new record and that holding otherwise would “render FOIA a nullity in the digital age.”¹³⁴ The case involved a FOIA request from the Center for Investigative Reporting (“CIR”) to the Bureau of Alcohol, Tobacco, Firearms and Explosives (“ATF”) for the “[t]otal number of weapons traced back to former law enforcement

¹²⁴ VERMONT STAT. ANN. §315 (2017).

¹²⁵ *Toensing v. AG of Vt.*, 2017 VT 99, ¶ 26, 178 A.3d 1000, 1009-10 (Vt. 2017).

¹²⁶ Email from Gary Derr, University of Vermont, to Jonathan Anderson (July 20, 2020) (on file with authors).

¹²⁷ *Appalachian Mountain Advocate, Inc. v. West Virginia Department of Environmental Protection*, Civ. Action No. 14-C-985 (Kanawha County Cir. Ct., Oct. 20, 2014) (Slip Opinion).

¹²⁸ Email from FOIA Officer, West Virginia University, to Jonathan Anderson (July 21, 2020) (on file with authors).

¹²⁹ NMSA 1978 § 14-2-8(B) (2009).

¹³⁰ HAW. REV. STAT. §92-1 et seq.; OIP Op. Ltr. No. 10-02.

¹³¹ NMSA 1978 § 14-3-15.1 and § 14-3-18 (2009).

¹³² Email from Amy Luke, University of Hawaii, to Sarah K. Wiley (July 21, 2020).

¹³³ Telephone conversation with Brandon Toensing, University of New Mexico (July 8, 2020) (notes on file with authors).

¹³⁴ *Ctr. for Investigative Reporting v. Dep’t of Justice*, 982 F.3d 668 (9th Cir. 2020).

ownership.”¹³⁵ Among the reasons that the ATF denied the request was that running a search of a database would constitute the creation of a new record, which is not required under FOIA.¹³⁶ The Ninth Circuit disagreed and remanded the case back to the District Court to assess whether the ATF can produce the data CIR requested.¹³⁷

The case is somewhat unusual in that ATF also argued that federal law specifically precluded disclosure of the underlying database, although the Ninth Circuit also rejected that claim.¹³⁸ The Ninth Circuit found that the ATF could either release the queried trace records, which would enable CIR to count them, or just disclose the statistical “count” that results from the query.¹³⁹ The ruling applied the 1996 EFOIA amendments that expanded the definition of “record” to include electronic information and the definition of “search” to include the review of computer records.¹⁴⁰ “In some ways, typing a query into a database is the modern day equivalent of physically searching through and locating data within documents in a filing cabinet,” the court wrote.¹⁴¹ “The subset of data selected is akin to a stack of redacted paper records.”¹⁴² The Ninth Circuit decision is the latest in a line of cases that have similarly held, based on EFOIA, that querying a database is not the creation of a new record.¹⁴³

The format of the data that are released is also important. As illustrated above, journalists are increasingly using computational methods to examine massive amounts of government data. As governmental data collection increases, such methods will become necessary to hold governments accountable without the need for news organizations to invest massive amounts of time and money. However, numerous responses to our requests for structured data resulted in the production of unstructured files or with a link to non-downloadable data, thus leaving the requester to expend valuable resources to transfer the data to the correct requested format.

Meanwhile, several universities responded with links to public websites that already contained information substantially similar to the spending data we were seeking. Although this proactive practice was helpful, the web pages often did not contain the exact information we requested and did not include a data download function. Such a practice highlights how data access is often strategic and offered by institutions with a specific outcome in mind.¹⁴⁴ Public agencies may proactively disclose limited data or post data in restrictive formats to control information. For example, the University of South Carolina has a “Spending Transparency” site that allows the public

¹³⁵ *Id.* at 6.

¹³⁶ *Id.*

¹³⁷ *Id.* at 42-44.

¹³⁸ *Id.* at 26.

¹³⁹ *Id.* at 42-43.

¹⁴⁰ *Id.* at 38-39.

¹⁴¹ *Id.* at 40.

¹⁴² *Id.*

¹⁴³ *See, e.g.,* Long v. Immigr. & Customs Enf’t, No. 17-CV-01097 (APM), 2018 WL 4680278, at *4 (D.D.C. Sept. 28, 2018) (“[N]either ‘sorting a pre-existing database of information to make information intelligible,’ nor ‘extracting and compiling data ... as to any discrete pieces of information that [an] agency does possess in its databases,’ amounts to the creation of a new agency record.”); Nat’l Sec. Couns. v. C.I.A., 898 F. Supp. 2d 233, 270 (D.D.C. 2012), *aff’d sub nom.* Nat’l Sec. Couns. v. Cent. Intel. Agency, 969 F.3d 406 (D.C. Cir. 2020) (“[S]orting a pre-existing database of information to make information intelligible does not involve the creation of a new record”); Schladetsch v. Dept of H.U.D., No. 99-0175, 2000 WL 33372125, at *3 (D.D.C. Apr. 4, 2000) (“Because HUD has conceded that it possesses in its databases the discrete pieces of information which [plaintiff] seeks, extracting and compiling that data does not amount to the creation of a new record.”). *See also* Baranetsky, *supra* note 9.

¹⁴⁴ Sergio Splendore, *Quantitatively Oriented Forms of Journalism and their Epistemology*, 10 SOCIO. COMPASS 343, 352 (2016).

to view vendor payments.¹⁴⁵ However, users must confine their searches to a particular fiscal year, month, campus, category, and sometimes multiple levels of subcategories before vendors and payment amounts are listed. This clunky and slow search process inhibits the ability of the press and public to fully scrutinize how the university is spending money. In South Carolina's response letter, the records custodian wrote that if we wanted more information about particular payments or vendors, we should let them know and they would "work on the cost to get you the more detailed information."¹⁴⁶ For members of the public and reporters on deadline, South Carolina's restrictive spending site might be sufficient. But there is no reason why the university also cannot proactively disclose the underlying dataset for the spending website.

Speculations on the divergent range of responses

What accounts for the divergent range of responses from universities? We suspect that several factors may be at play. First, differences in legal requirements among states may have shaped how universities responded. As noted above, some of the universities that have responded comparatively earlier are in states that require some sort of response within 10 business days. Other legal factors may include the particular law in each state governing how agencies must process requests for data runs and what types of fees they can impose.

Another potential factor affecting access is the technological knowledge and capabilities of universities. Each university has different policies and procedures for how to respond to records requests. It is possible that universities that provide greater access have more robust processes to extract data from databases or have more experienced and knowledgeable information technology staff.

Finally, a third factor to consider is that universities that responded better may have more employees whose job duties entail responding to public records requests. Some universities have offices or employees focused solely on public records duties, while other universities assign that task to communications, legal, finance, or administrative employees.

Recommendations

Based on the results of the study, we propose several ideas to improve access to structured data held by the government.

- First, we suggest that states revise their public records laws to expressly declare that querying a database is a duty of a records custodian and is not the creation of a new record. Codifying this in statute would make clear that the public has a right to information stored in a database, even if accessing that information may be somewhat complex, or at least potentially more so than copying a piece of paper.
- Second, we suggest that when agencies procure new information systems and software, they ensure that the underlying data are retained and can be accessed easily in response to public records requests. This change can take the form of voluntary internal policy, local ordinance, or state administrative rule or statute. Following this

¹⁴⁵ Financial Transparency Reporting, Office of the Comptroller, University of South Carolina, https://www.sc.edu/about/offices_and_divisions/controller/financial_reporting/financial_transparency_reporting/index.php.

¹⁴⁶ Email from Edwin E. Evans, University of South Carolina, to Jonathan Anderson (July 23, 2020) (on file with authors).

step would help prevent access hurdles premised on arguments that data do not exist or would be expensive to produce.

- Third, states should require that agencies proactively disclose the existence of databases that they use, similar to the requirement in California. Such disclosures should also contain a description of the fields of the databases. Disclosing such information would help record requesters make more targeted record requests, which would also save agencies time and effort in processing such requests.
- Fourth, to the extent that agencies proactively post data online, they should make the whole dataset available in a common file format, such as CSV or Excel, and not make analysis prohibitively complicated or time consuming like the University of South Carolina's spending website.
- Fifth, we suggest that if record requesters are met with burdensome fees or claims that querying data would require the creation of a new record, that they inquire with a record custodian about the particular processes that would be required to produce the data. It is apparent from these findings that some agencies may not understand how to find certain types of data or produce it in a structured format.

Conclusion

As governmental operations at all levels have become increasingly digitized, records of those activities have also moved from paper and filing cabinets to databases and servers. The research reported in this article suggests that units of government—here, public universities—have the capacity to produce structured data containing common administrative information reasonably quickly and with little to no cost to a requester. However, this was far from the norm. The universities in our sample produced responsive structured data without a fee in slightly more than a quarter of requests, meaning the vast majority of our requests failed to yield the information we sought in a structured data format and for free. Some requests were met with implausible claims that records did not exist or would cost hundreds of dollars or more than \$1,000 to produce. Some universities' assertions that querying the requested data would require creation of a new record appear to be inconsistent with case law. Indeed, the results indicate that agency responses are highly variable, and we speculate that variance may be dependent, in part, on jurisdictional legal requirements, agency technological knowledge and capabilities, and staffing levels of offices tasked with processing records requests.

More research is needed to better understand why agencies responded to the study's records requests differently and whether factors that improved access can be replicated elsewhere. Future research should entail auditing more government agencies and increasing the sample size. It would also be valuable to update analyses of state and federal court decisions about public access to databases, as much of that scholarship was published decades ago before key legislative changes and court decisions.

To be sure, this paper is not without limitations. Perhaps due in part to the COVID-19 pandemic, many responses were slow to arrive, and several universities gave notice that extra processing time was needed. Another limitation relates to how we made the requests. As explained above, three distinct items were requested in one correspondence. Although we felt it would be beneficial to ask for a variety of data, combining the requests may have skewed results if one database was more difficult to access or query than another, potentially lengthening the overall response time or increasing the fee estimate for all records. It may be more prudent to limit requests to one item per

agency in future studies. As we focused on higher educational institutions, the results may differ with other agencies and at other levels of government.

Still, the findings of this study offer actionable insights. For policymakers, we suggest limiting the ability of agencies to claim that querying a database constitutes creation of a new record. Agencies can help improve access to structured data by posting on their websites information about their databases, making datasets easily accessible online in a structured and downloadable format, and ensuring that the information systems they use can easily query and produce structured data. Finally, for record requesters, we suggest investigating the precise technological capabilities of agencies when access problems arise.

Appendix A

Example of public records requests filed with universities

[Date]

VIA ONLINE PORTAL

Rutgers University, New Brunswick
University Ethics and Compliance
<https://uec.rutgers.edu/programs/opra/>

Re: Public Records Request

To Whom It May Concern:

Pursuant to the state public records law, N.J. Stat. Ann. Secs. 47:1-1 to 47:1A-13, I write to request the following public records:

1. All entries in any accounting or bookkeeping databases (i.e., software) showing all payments that Rutgers University-New Brunswick made to vendors between January 1, 2020 and March 31, 2020, including but not limited to the following fields to the extent they are available:
 - a. Name of vendor
 - b. Amount of payment
 - c. Date of payment
 - d. Invoice and/or contract number
 - e. Purpose of expense
2. All entries in any payroll, accounting, or computerized bookkeeping databases (i.e., software) showing all payroll payments made to Rutgers University-New Brunswick employees in 2019, including but not limited to the following fields to the extent they are available:
 - a. Name of employee
 - b. Title of employee
 - c. Classification or rank of employee
 - d. Amount of payroll payment
 - e. Date of payment
3. An email log of all emails Chancellor Christopher J. Molloy sent and/or received between March 9-15, 2020, including but not limited to the following fields to the extent they are available:
 - a. Date of email
 - b. Time of email
 - c. Name of sender ("FROM" field)
 - d. Sender email address
 - e. Name of recipient ("TO" field)
 - f. Recipient email address

- g. Name of CC recipient (“CC” field)
- h. CC recipient email address
- i. Name of BCC recipient (“BCC” field)
- j. BCC recipient email address
- k. Subject line

For each of these three requests, I request these data as an Excel spreadsheet, comma-delimited text file, or if neither formats are possible, the existing format of the responsive data.

The listed fields may not be the exact name of the fields in the databases, but rather describe the nature of the information contained in the requested fields.

Please notify me if the total cost of this request will exceed \$25. If the total cost of this request exceeds \$25, written authorization from me will be necessary to proceed with the processing of this request.

If you deny my request, or even part of my request, please do so in writing and state the reason for your denial.

If you have any questions or wish to communicate with me about this request, please email me at [email address] or call me at [phone number]. Thank you.

Sincerely,

/s/

[Name of Requester]

Appendix B

Table of substantive responses received by March 5, 2021

1=Produced Responsive Data

2=Produced/Offered Different Data

3=No Responsive Records

4=Required Prepayment Before Processing

5=Required In-Person Inspection

6=No Substantive Response Yet

7=Denied

*=Received Reminder Inquiry

University	Vendor Data	Payroll Data	Email Log
Arizona State University, Tempe*	6	6	6
Boise State University	1	1	1
College of Southern Nevada	1	1	1
Indiana University, Bloomington	1	1	3
Iowa State University	2	2	1
Louisiana State University	1	5	6
Michigan State University	4	1	3
Montana State University	1	1	3
North Carolina State University at Raleigh*	1	3	3
North Dakota State University	2	6	6
Ohio State University*	1	2	3
Oregon State University	1	6	6
Rutgers University, New Brunswick	1	7	4
South Dakota State University	1	2	6
Texas A & M, College Station	1	1	1
University at Buffalo	1	1	3
University of Alaska, Anchorage*	6	6	6
University of California, Los Angeles	3	3	3
University of Central Florida	1	4	4
University of Colorado, Boulder*	2	2	4

University of Connecticut	1	1	1
University of Hawaii at Mona	3	3	3
University of Illinois at Urbana-Champaign	1	1	1
University of Kansas	4	4	4
University of Kentucky*	6	6	7
University of Maine, Orono*	1	2	3
University of Maryland, University College*	6	6	6
University of Massachusetts, Amherst	4	4	4
University of Minnesota, Twin Cities*	1	1	4
University of Mississippi*	2	4	3
University of Missouri, Columbia*	3	3	3
University of Nebraska, Lincoln	3	3	3
University of New Hampshire, Main Campus*	1	2	1
University of New Mexico, Main Campus	1	2	3
University of Oklahoma, Norman Campus*	6	6	6
University of Rhode Island	4	4	4
University of South Carolina, Columbia	2	2	4
University of Vermont, Burlington	3	3	3
University of Washington, Seattle Campus	1	1	6
University of Wisconsin, Madison*	4	4	6
University of Wyoming, Laramie	4	4	4
Utah Valley University	1	2	4
West Chester University of Pennsylvania	1	1	1
West Virginia University	2	2	3