

No. 18-17356

**IN THE UNITED STATES COURT OF APPEALS
FOR THE NINTH CIRCUIT**

THE CENTER FOR INVESTIGATIVE REPORTING,
Plaintiff-Appellant,

v.

UNITED STATES DEPARTMENT OF JUSTICE,
Defendant-Appellee.

*On Appeal from the United States District Court
for the Northern District of California, San Francisco
No. 3:17-cv-06557-JSC,
The Honorable Jacqueline Scott Corley, United States Magistrate Judge*

**BRIEF OF AMICI CURIAE FIVE MEDIA ORGANIZATIONS AND
SIXTEEN DATA JOURNALISTS IN SUPPORT OF APPELLANT AND
REVERSAL**

Dated: March 28, 2019

Mason A. Kortz
Kendra K. Albert
Cyberlaw Clinic
Harvard Law School
1585 Massachusetts Ave.
Cambridge, MA 02139
(617) 495-2845
mkortz@law.harvard.edu

Counsel for *amici curiae*

CORPORATE DISCLOSURE STATEMENT

Amici curiae Investigative Reporters and Editors Inc.; The Media Law Resource Center, Inc.; The MuckRock Foundation; Freedom of the Press Foundation; and The Reporters Committee for Freedom of the Press have no parent corporations. They have no stock, and therefore, no publicly held company owns 10% or more of their stock.

Dated: March 28, 2019

s/ Mason A. Kortz

Mason A. Kortz
Cyberlaw Clinic
Harvard Law School
1585 Massachusetts Ave.
Cambridge, MA 02139
(617) 495-2845
mkortz@law.harvard.edu

STATEMENT OF COMPLIANCE WITH RULE 29

Pursuant to Federal Rule of Appellate Procedure 29(a)(2), *amici curiae* certify that all parties have consented to the filing of this brief.

Pursuant to Federal Rule of Appellate Procedure 29(a)(4)(E), *amici curiae* certify that no party's counsel authored this brief in whole or in part; no party or party's counsel contributed money that was intended to fund preparing or submitting this brief; and no person—other than the *amici curiae*, their members, or their counsel—contributed money that was intended to fund preparing or submitting this brief.

TABLE OF CONTENTS

CORPORATE DISCLOSURE STATEMENT	i
STATEMENT OF COMPLIANCE WITH RULE 29.....	ii
TABLE OF CONTENTS.....	iii
TABLE OF AUTHORITIES	iv
STATEMENT OF INTEREST OF AMICUS CURIAE	1
SUMMARY OF ARGUMENT	9
ARGUMENT	10
I. The “new records” doctrine should be applied to databases in a context-sensitive manner.....	12
A. Databases have distinct characteristics that are relevant to the application of the “new records” doctrine.	12
B. The “new records” doctrine arose before the widespread use of databases.	16
II. Correct application of the “new records” doctrine would treat database queries as accessing existing records.....	18
A. Courts have failed to meaningfully distinguish between new records and existing records in the database context.....	19
B. Accessing a particular arrangement of information by querying a database does not amount to creation of a new record.	21
C. Treating queries as accessing existing records avoids absurd results. ..	25
III. Treating queries as accessing existing records does not inappropriately expand the obligations of agencies under FOIA.....	28
CONCLUSION	30
CERTIFICATE OF COMPLIANCE.....	31
CERTIFICATE OF SERVICE	32

TABLE OF AUTHORITIES

Cases

<i>Am. Civil Liberties Union v. Arizona Dep’t of Child Safety</i> , 240 Ariz. 142 (Ct. App. 2016)	21
<i>Ctr. for Investigative Reporting v. U.S. Dep’t of Justice</i> , No. 17-CV-06557-JSC, 2018 WL 3368884 (N.D. Cal. July 10, 2018)	11
<i>Forsham v. Califano</i> , 587 F.2d 1128 (D.C. Cir. 1978)	16
<i>Forsham v. Harris</i> , 445 U.S. 169 (1980)	16, 22
<i>Johnson v. CIA</i> , 330 F. Supp. 3d 628 (D. Mass. 2018)	20
<i>Kissinger v. Reporters Comm. for Freedom of the Press</i> , 445 U.S. 136 (1980)	21, 22
<i>Landmark Legal Found. v. U.S. Dep’t of Justice</i> , 211 F. Supp. 3d 311 (D.D.C. 2016)	21
<i>Long v. U.S. Dep’t of Justice</i> , 450 F. Supp. 2d 42 (D.D.C. 2006), <i>order amended on reconsideration</i> , 457 F. Supp. 2d 30 (D.D.C. 2006), <i>amended</i> , 479 F. Supp. 2d 23 (D.D.C. 2007)	14
<i>Nat’l Sec. Counselors v. CIA</i> , 898 F. Supp. 2d 233 (D.D.C. 2012)	18, 19, 20, 25, 26, 27, 29
<i>NLRB v. Sears, Roebuck & Co.</i> , 421 U.S. 132 (1975)	16, 22
<i>People for American Way Foundation v. U.S. Dep’t of Justice</i> , 451 F. Supp. 2d 6 (D.D.C. 2006)	19, 20, 26, 29
<i>Schladetsch v. U.S. Dep’t of HUD</i> , No. 99-0175, 2000 WL 33372125 (D.D.C. Apr. 4, 2000)	18, 22
<i>Tereshchuk v. Bureau of Prisons</i> , 67 F. Supp. 3d 441 (D.D.C. 2014), <i>aff’d sub nom. Tereshchuk v. Bureau of</i> <i>Prisons, Dir.</i> , No. 14-5278, 2015 WL 4072055 (D.C. Cir. June 29, 2015)	21, 22
<i>Thompson Pub. Grp., Inc. v. Health Care Fin. Admin.</i> , No. CIV. A. 92-2431-LFO, 1994 WL 116141 (D.D.C. Mar. 15, 1994)	13, 26, 28

U.S. Dep’t of State v. Ray,
502 U.S. 164 (1991).....26

Yeager v. DEA,
678 F.2d 315 (D.C. Cir. 1982)..... 12, 16, 17, 21, 28

Statutes

5 U.S.C. § 552..... 11, 17

E-FOIA Amendments of 1996,
Pub. L. No. 104-231, 110 Stat. 3048, 3049 (1996).....17

Tiahrt Amendment,
Pub. L. No. 112-55, 125 Stat. 552, 609-10 (2011)11

Other Authorities

Daniel Simmons-Ritchie, *Still Failing the Frail: The Data and Records Behind Our Reporting and Our Database*, PennLive (Oct. 14, 2018)10

Get The Data: Explore Data on All Police Shootings From the Nation’s 50 Largest Local Police Departments, VICE News (Dec. 10, 2017)10

J.K. Rowling, *Harry Potter and the Order of Phoenix* (First American Ed. 2003) 16

John Foley, *ATF’s Gun Tracing System is a Dud*, InformationWeek (Mar. 13 2013, 12:12 PM)14

Office of Information Policy, *Defining a “Record” Under the FOIA*, U.S. Dep’t of Justice (last updated Feb. 15, 2017) 23, 24, 25

P. Stephen Gidiere III, *The Federal Information Manual* (2013).....18

RAND Corporation, *Strategies for Disrupting Illegal Firearms Markets*13

Robert Benincasa, *Search The Thousands Of Disaster Buyouts FEMA Didn't Want You To See*, NPR (Mar. 5, 2019).....10

Legislative Materials

H.R. Rep. No. 104-795 (1996)..... 17, 18

STATEMENT OF INTEREST OF AMICUS CURIAE

Amici curiae consist of five media and media-related organizations and sixteen individual data journalists and professors of journalism. Collectively, *amici* have deep experience with the technical aspects of working with structured data, including databases, and with government transparency processes, including the Freedom of Information Act. As such, *amici* have a significant interest in a strong right of access to records held in government databases. Brief statements of the expertise of organizational and individual *amici* follow.¹

Investigative Reporters and Editors Inc. (“IRE”) is a grassroots non-profit organization dedicated to improving the quality of investigative reporting. IRE was formed in 1975 to create a forum in which journalists throughout the world could help each other by sharing story ideas, newsgathering techniques, and news sources. IRE provides members access to thousands of reporting tip sheets and other materials through its resource center and hosts conferences and specialized training throughout the country.

The Media Law Resource Center, Inc. (“MLRC”) is a non-profit professional association for content providers in all media, and for their defense lawyers, providing a wide range of resources on media law and policy issues. These include news and analysis of legal, legislative, and regulatory developments;

¹ Biographies of individual *amici* are provided solely for identification purposes.

litigation resources and practice guides; and national and international media law conferences and meetings. The MLRC also works with its membership to respond to legislative and policy proposals and speaks to the press and public on media law and First Amendment issues. It counts as members over 130 media companies, including newspaper, magazine, and book publishers, TV and radio broadcasters, and digital platforms, and over 200 law firms working in the media law field.

The MuckRock Foundation is a journalism and government transparency non-profit that has helped thousands of requesters around the United States better file, share, and understand Freedom of Information requests. This work has often involved obtaining and analyzing federal databases, including data on the government's 1033 program that led to reforms of this program. They often work with agency FOIA personnel and IT departments to help craft requests for data that protects privacy and reduces the burden on agency staff while providing key insights into government operations.

Freedom of the Press Foundation is a non-profit organization that supports and defends public interest journalism focused on transparency and accountability. The organization works to preserve and strengthen rights guaranteed to the press through a variety of avenues, including public advocacy, legal advocacy, the promotion of digital security tools, and crowdfunding.

The Reporters Committee for Freedom of the Press is an unincorporated non-profit association. The Reporters Committee was founded by leading journalists and media lawyers in 1970 when the nation's news media faced an unprecedented wave of government subpoenas forcing reporters to name confidential sources. Today, its attorneys provide *pro bono* legal representation, *amicus curiae* support, and other legal resources to protect First Amendment freedoms and the newsgathering rights of journalists.

Dana Amihere is the data editor at Southern California Public Radio (KPCC). She's a designer, developer, and data journalist who has previously worked for The Dallas Morning News, The Baltimore Sun, and Pew Research Center. She has worked with government databases to report on education, homelessness, and inequality in the criminal justice system. She is currently helping manage data for a statewide news collaborative, the California Reporting Project, which aims to uncover police misconduct through records obtained under the state's new transparency law.

Zita Arocha is an Associate Professor of Practice in Journalism at the University of Texas-El Paso. She is the director and founder of Borderzine.com, a multimedia web magazine at UTEP that prepares Hispanic college journalists for jobs in 21st-century newsrooms. Before joining UTEP, she previously worked as a reporter over 20 years for The Washington Post, Miami Herald, Miami Times, and

Tampa Times. She was Executive Director of the National Association of Hispanic Journalists from 1993-1997.

Tim Broderick is the data reporter for Modern Healthcare magazine. Previously he was the Editor of Graphics and Data Journalism with the Daily Herald, which covers the suburbs surrounding Chicago. He is a graduate of the Eastern Illinois University journalism program and has numerous SPJ Lisagor and Illinois AP Editors awards, including for his part in an influential series on low-income students and test scores in Illinois.

Carli Brosseau is a data reporter at The News & Observer in Raleigh and a longtime member of Investigative Reporters and Editors. She earlier worked as a data reporter at The Oregonian and the Arizona Daily Star. Carli has advocated for easier access to data under state public records laws, including proposing additions to Oregon's procurement statutes that were passed into law. Her reporting on criminal justice issues has received state and national recognition.

Meredith Broussard is an Assistant Professor at the NYU Arthur L. Carter Journalism Institute. She is the author of *Artificial Unintelligence: How Computers Misunderstand the World* (MIT Press 2018) and "Big Data in Practice: Enabling Computational Journalism Through Code-Sharing and Reproducible Research Methods." Her research focuses on the use of data analysis in investigative reporting.

Matt Carroll is a former investigative reporter with the Boston Globe. He worked on the Globe's Spotlight team and, along with other members of the team, was a winner of the 2003 Pulitzer Prize for Public Service for its coverage of the Catholic church sexual abuse crisis.

Stephen K. Doig is a professor of data journalism at the Cronkite School of Journalism of Arizona State University. Before joining ASU in 1996, he was an Associate Editor/Researcher of the Miami Herald, where he worked for 20 years. Various data projects on which he worked at The Herald and at ASU have won the Pulitzer Prize for Public Service, the IRE Award, the Goldsmith Prize for Investigative Reporting, the George Polk Award, and other recognition.

Dan Keating has been at The Washington Post for 20 years doing stories on politics, health, crime, demographics, education, procurement, and other topics. Previously, he worked as a beat reporter at The Berkshire Eagle in Pittsfield, Mass. and The Miami Herald in North Miami, Fla., and Key West, Fla., before specializing in data analysis for stories and graphics. He shared the Pulitzer Prize for Investigative Reporting in 1999.

Dhrumil Mehta is a database journalist at FiveThirtyEight and an adjunct lecturer in Public Policy at the Harvard Kennedy School. He builds and maintains databases related to U.S. politics and elections, including FiveThirtyEight's database of public opinion polls. He has also designed and teaches a course called

“Programming and Data for Policymakers” at the Harvard Kennedy School intended to teach technology skills to students in the School of Government.

Miguel Paz is a journalist and digital media strategist and Distinguished Lecturer in Data Journalism at the Craig Newmark Graduate School of Journalism at CUNY and a former Tow Professor. Previously, he was a Nieman Berkman Fellow at Harvard University and a Knight International Journalism Fellow at the International Center for Journalists.

Cheryl Phillips is the Lorry I. Lokey Visiting Professor in Professional Journalism at Stanford University's Department of Communication and director of the Big Local News project, part of Stanford's Journalism and Democracy Initiative. She also is a co-founder of the Stanford Open Policing Project and involved in the California Civic Data Coalition. Previously, she worked at the Seattle Times as a Data Innovation Editor, where she was twice involved in coverage of breaking news that resulted in Pulitzer Prizes.

Jeff South has been an associate professor in the Robertson School of Media and Culture at Virginia Commonwealth University since 1997. Before moving into academia, he was a newspaper reporter and editor for 20 years in Texas, Arizona, and Virginia. South teaches data journalism and other courses. He has won teaching awards from VCU and the Society of Professional Journalists as well as a Fulbright.

Samantha Sunne is a freelance journalist based in New Orleans, Louisiana. She is the recipient of three national grants for investigative reporting and a frequent speaker at conferences, universities, and newsrooms around the world, teaching digital tools and tech for journalists. She founded a meetup group for journalists and technologists to promote open data and open government in New Orleans.

Serdar Tungoren teaches data journalism at the Stanford University Graduate Journalism Program. Prior to joining Stanford, he worked for The Associated Press, The Washington Post, Congressional Quarterly, and local news outlets in California, Connecticut, and New Jersey.

Lucia Walinchus is an award-winning journalist, attorney, and author. A 2016 Fulbright Berlin Scholar, she has been featured as a guest speaker on CNN and is a contracted freelancer for the New York Times. Last year, Walinchus started as the new executive director at the Ohio Center for Investigative Journalism, Eye on Ohio.

Fedor Zarkhin is a data reporter for The Oregonian/OregonLive, a daily newspaper in Portland, Oregon. He has produced impactful projects based on analyses of elder abuse data, environmental enforcement data, and residential demolition data. He regularly uses programming languages to query data, including Structured Query Language and Python. Zarkhin has been honored with a variety of

journalism awards, including the 2018 National Headliner Award in healthcare reporting.

SUMMARY OF ARGUMENT

This is a case about how far the right of the public to know what their government is up to extends in a digital age where agencies increasingly use electronic formats rather than filing cabinets to store records. Failure to create a context-sensitive rule that acknowledges these differences could further muddy the waters about how exactly courts should navigate the Freedom of Information Act in the context of databases.

Electronic databases are different from other methods of data storage in that they store information in a highly structured format, designed to be queried in many ways. Although information in a database can be extracted in an almost limitless number of arrangements, each of these arrangements involves access to the same, existing data. Despite this, some courts have applied the “new records” doctrine, which was developed before the advent of electronic record-keeping, to limit access to information contained within databases.

The lower court here held that querying a database for aggregate data entails the creation of a new record. But the case law that the lower court relied on fails to distinguish new records and existing records meaningfully in the context of databases. A better approach is to treat any simple database query or series of queries as accessing existing records. This rule is consistent with both congressional intent and general principles of FOIA jurisprudence. Moreover, it avoids absurd results

that arise from drawing arbitrary distinctions between new and existing records in the database context and would not unduly burden agencies. *Amici*, therefore, urge this Court to reverse the lower court’s decision, and draw a new line in the “new records” doctrine that better conforms to the way agencies actually use electronic databases.

ARGUMENT

Access to database totals may not, on its face, seem like the sexiest topic, even within the Freedom of Information Act. But database journalism is now fundamental to modern newsrooms, leading to impactful stories on police violence, disaster buyouts, and nursing home abuse.² The nuances of exactly how journalists can request and use information from databases, governed by rules like the “new records” doctrine, has the potential to make or break efforts to hold the government accountable using its own data.

² *Get The Data: Explore Data on All Police Shootings From the Nation’s 50 Largest Local Police Departments*, VICE News (Dec. 10, 2017), https://news.vice.com/en_us/article/a3jjpa/nonfatal-police-shootings-data [<https://perma.cc/AZ2H-A6PJ>]; Robert Benincasa, *Search The Thousands Of Disaster Buyouts FEMA Didn't Want You To See*, NPR (Mar. 5, 2019), <https://www.npr.org/2019/03/05/696995788/search-the-thousands-of-disaster-buyouts-fema-didnt-want-you-to-see> [<https://perma.cc/343U-DSTS>]; Daniel Simmons-Ritchie, *Still Failing the Frail: The Data and Records Behind Our Reporting and Our Database*, PennLive (Oct. 14, 2018), https://www.pennlive.com/midstate/2018/10/still_failing_the_frail_the_da.html [<https://perma.cc/5K7S-VC4T>].

The unique facts of this case make it the ideal vehicle for adoption of a bright-line rule for the “new records” doctrine. Ordinarily, a person requesting public information through the Freedom of Information Act (“FOIA”), 5 U.S.C. § 552, could request non-aggregate data, run their own analysis, and avoid the “new records” doctrine entirely. However, below, the district court held that non-aggregate data from the federal Firearms Trace System (“FTS”) database could not be disclosed under the Tiahrt Amendment, Pub. L. No. 112-55, 125 Stat. 552, 609-10 (2011). *See Ctr. for Investigative Reporting v. U.S. Dep’t of Justice*, No. 17-CV-06557-JSC, 2018 WL 3368884, at *9 (N.D. Cal. July 10, 2018). Since aggregate data from the FTS database could be released under the Tiahrt Amendment, the validity of the request at issue here—as interpreted by the district court—turns on whether a request for “aggregate data” constitutes the creation of a new record.

The district court ultimately held that the extraction of “aggregate data” from the FTS database would constitute the creation of a new record under FOIA. *See id.* at *9–10. In doing so, the court drew a distinction between aggregate and non-aggregate data that is meaningless with regards to database structure and difficult to implement in the real world. If the court arrives at this question, *amici* respectfully urge the court to hold that any arrangement of information—that is, any record—that can be retrieved from a database with a simple query or series of queries

constitutes an existing record for FOIA purposes. Adopting this rule would set forth a clear and workable standard in an otherwise murky area of law.

I. The “new records” doctrine should be applied to databases in a context-sensitive manner.

In concluding that the requesters were not entitled to aggregate data from the FTS database, the lower court applied the “new records” doctrine, which limits agencies’ obligations to create new records in response to FOIA requests. This doctrine was originally developed in the age of paper and filing cabinets and only later extended to electronic records. As a result, courts must take care when applying the doctrine to modern records storage systems, especially databases.

A. Databases have distinct characteristics that are relevant to the application of the “new records” doctrine.

Records stored in databases are subject to FOIA, just like other agency records. *See Yeager v. Drug Enf’t Admin.*, 678 F.2d 315, 322 (D.C. Cir. 1982). In some ways, though, databases behave differently than physical records or even other electronic records. Consider the classic example of paper records stored in a filing cabinet. A record is usually a document that contains information in a fixed form. That information cannot be restructured without expending time and effort. Likewise, each record is sorted into a single folder, and those folders are organized in a specific order. There is only one arrangement of records at a given time, and creating a new arrangement takes time and effort.

Databases are different, in that they are designed to provide access to information in multiple different arrangements with minimal effort. Databases have this functionality because they store information in a highly structured, easily searchable form. As the United States District Court for the District of Columbia has explained,

A traditional FOIA search could include physically retrieving 1000 different paper forms or “records,” each of which has been placed in a separate file, and each of which contains only one relevant paragraph. This would be analogous to (and much more difficult than) a computer “query search” for those paragraphs. If defendant has categorized its data by size of employer, for example, and the plaintiff’s request can be retrieved by a single search or a simple series of searches, then the information exists in the form of parts of multiple “records.”

Thompson Pub. Grp., Inc. v. Health Care Fin. Admin., No. CIV. A. 92-2431-LFO, 1994 WL 116141, at *2 (D.D.C. Mar. 15, 1994). The entire point of a database is to bring these separate pieces of information together in different arrangements. Extracting these arrangements does not require any information to be added to the database—it only requires reading existing structural information.

To make this point clear, it is helpful to understand how databases store information. The FTS database at issue in the present case is, to the best of *amici*’s knowledge, built in Oracle, a relational database management system.³ In a typical

³ See RAND Corporation, *Strategies for Disrupting Illegal Firearms Markets*, 15 n.3 (2008), available at https://www.rand.org/content/dam/rand/pubs/technical_reports/2008/RAND_TR512.pdf [<https://perma.cc/N46A-KK8U>] (“Since 1998, FTS has run on an Oracle® database and application platform”); John

relational database, information is organized into tables, which are composed of rows and columns. The set of tables or other data structures that store information in a database are often referred to as the *database schema*. The database schema is sometimes referred to as the *physical model*, because it describes how information is broken down into discrete segments for physical storage on a computer.

The physical model of a database will often be different than the conceptual organization of the information stored in that database. Relational database management systems allow related information to be spread across multiple tables. *See Long v. U.S. Dep't of Justice*, 450 F. Supp. 2d 42, 48 (D.D.C. 2006), *order amended on reconsideration*, 457 F. Supp. 2d 30 (D.D.C. 2006), *amended*, 479 F. Supp. 2d 23 (D.D.C. 2007) (“In a relational database, information is subdivided into tables, and connections or linkages can be programmed between the tables, allowing the data maintained in one table to be related to the data stored in another.”) The process of separating related information into multiple tables is called *normalization* and is an important practice that helps prevent errors from being introduced into the database. If a database is normalized, accessing a single conceptual “unit” of information may require retrieving information from multiple tables at once. What

Foley, *ATF's Gun Tracing System is a Dud*, InformationWeek (Mar. 13 2013, 12:12 PM), <https://www.informationweek.com/applications/atfs-gun-tracing-system-is-a-dud/d/d-id/1109062> [<https://perma.cc/Z4US-2SVN>] (“[FTS] pulls together information from a variety of sources, mostly Oracle databases.”).

this means is that a single conceptual arrangement of information—for example, an individual firearm trace—may never exist as a single discrete entity in the physical model of the database. Any one database can support a multitude of different arrangements of information.

Recreating a particular arrangement of information from the physical model of a database requires extracting related information and putting it together in a way that makes conceptual sense. This is done by running a *query* on the database. A query is an instruction that tells a database management system to select a specific subset of information from a database and return it in a particular arrangement.⁴ For example, in a relational database, a query specifies which rows and columns to extract from which tables, how they should be sorted, and whether they should be summed, averaged, counted, or otherwise manipulated. If the database is normalized, a query will also specify whether multiple tables should be joined together.

Returning to the analogy of papers in a filing cabinet, then, a database is like no filing cabinet that has ever existed. It is like a filing cabinet in which each document is broken down to the level of sentences and cross-referenced with every

⁴ There are other types of queries that instruct a database management system to insert, update, or delete data from a database, or to alter the structure of the tables, rows, columns, or other data structures that make up the database. Such queries are beyond the scope of FOIA, as they would require agencies to create, alter, or destroy records. For the purpose of this brief, *amici* use the term “query” to refer exclusively to queries that select information without changing the structure or contents of the database.

other related document. Moreover, in order to open the filing cabinet, one must first specify which arrangement the documents should be presented in—at which point, the sentences will be rearranged. Indeed, databases may have more in common with the famed Room of Requirement at Hogwarts⁵ than with how documents were stored on paper—except that databases are real, and have very real implications for access to records under FOIA.

B. The “new records” doctrine arose before the widespread use of databases.

Not found in the text of FOIA itself, the Supreme Court first articulated the “new records” doctrine in *NLRB v. Sears, Roebuck & Co.* in 1975, 421 U.S. 132, 162 (1975), stating that “[FOIA] only requires disclosure of certain documents which the law requires the agency to prepare or which the agency has decided for its own reasons to create.” The Court did not expand on the rationale behind the exception, *id.* at 161–62, and neither did subsequent cases, *see, e.g., Forsham v. Harris*, 445 U.S. 169, 186 (1980) (discussing *Sears*’s “new document” doctrine without speaking to the rationale); *Forsham v. Califano*, 587 F.2d 1128, 1136 (D.C. Cir. 1978) (applying the *Sears* test without further reasoning).

Seven years later in *Yeager v. DEA*, the D.C. Circuit read FOIA to cover electronic databases. *See* 678 F.2d 315, 321 (D.C. Cir. 1982). The court applied the

⁵ J.K. Rowling, *Harry Potter and the Order of Phoenix* 386–87 (First American Ed. 2003).

“new record” doctrine to electronic databases as well, regarding it as “well-settled.” *Id.* It noted that “[a]lthough accessing information from computers may involve a somewhat different process than locating and retrieving manually-stored records, these differences may not be used to circumvent the full disclosure policies of the FOIA.” *Id.* The court nevertheless held that “compact[ing]” a database by deleting exempt materials would constitute creation of a new record. *Id.* at 322–23. In doing so, it relied on the assumption that Congress had not “intended any manipulation or restructuring of the substantive content of a record when it commanded agencies to ‘delete’ exempt information.” *Id.* at 323.

This changed when Congress formally expanded FOIA to include records in “electronic format[s]” in the E-FOIA Amendments of 1996. Pub. L. No. 104-231, § 3, 110 Stat. 3048, 3049. In doing so, it acknowledged electronic databases as having a distinct structure in the FOIA context. *See* H.R. Rep. No. 104-795, at 22 (1996) (“Computer records found in a database rather than in a file cabinet may require the application of codes or some form of programming to retrieve the information.”). Congress further explained that in light of the requirement for agencies to “make reasonable efforts to search for the records in electronic form or format,” 5 U.S.C. § 552(a)(3), a search of an electronic database “would not amount to the creation of records.” H.R. Rep. No. 104-795, at 22 (1996). Congress specifically noted that if an electronic search were to constitute creating a new record,

“it would be virtually impossible to get records maintained completely in an electronic format, like computer database information, because some manipulation of the information likely would be necessary to search the records.” *Id.*

Court decisions about database-related FOIA requests have followed and expanded on Congress’ view. *See* P. Stephen Gidiere III, *The Federal Information Manual* § 6.3 (2013) (“[C]ourts . . . have rejected [the view that automated searches are creating new records] and required agencies to conduct automated searches of their electronic databases.”). In interpreting the E-FOIA Amendments, courts have held that “[t]he programming necessary to conduct [a] search [of a database] is a search tool and not the creation of a new record.” *Schladetsch v. U.S. Dep’t of HUD*, No. 99-0175, 2000 WL 33372125, at *3 (D.D.C. Apr. 4, 2000). More recently, the United States District Court for the District of Columbia relied on the rationale behind the E-FOIA Amendments to hold that “sorting a pre-existing database of information to make information intelligible does not involve the creation of a new record.” *Nat’l Sec. Counselors v. CIA*, 898 F. Supp. 2d 233, 270 (D.D.C. 2012).

II. Correct application of the “new records” doctrine would treat database queries as accessing existing records.

In the case below, the court applied the “new record” doctrine without careful consideration of the special characteristics of databases described above. As a result, it adopted a distinction that does not reflect the realities of modern record-keeping. *Amici* urge this Court to reverse that decision and hold that when a simple query or

set of queries is used to access information already available in a database, the resulting arrangement of data constitutes an existing record. This approach is most consistent with the E-FOIA Amendments, judicial interpretations of FOIA, and the Department of Justice's own guidance. Moreover, this approach avoids unworkable results that would otherwise arise if requesters, agencies, and courts were required to distinguish between queries that access existing records and queries that create new records.

A. Courts have failed to meaningfully distinguish between new records and existing records in the database context.

While there is consensus that searching and sorting data in a database does not constitute the creation of a new record, *see Nat'l Sec. Counselors*, 898 F. Supp. 2d at 270, courts have held that other database functions do. These distinctions are arbitrary and do not reflect the unique characteristics of databases.

For example, some courts have held that requests for a list or index of data in a database—as opposed to the data points themselves—constitute the creation of a new record and are thus not mandated by FOIA. The content-index distinction appears to have been first introduced in *People for American Way Foundation v. U.S. Department of Justice*, 451 F. Supp. 2d 6, 15 (D.D.C. 2006). In that case, the plaintiff requested records that could be located using the federal PACER database. *Id.* at 9. In dicta, the court explained what would have happened if the plaintiff had not asked for the records themselves, but had instead requested the list of results

from PACER. *Id.* at 15. Noting that PACER was an “outside database,” the court stated that “if plaintiff sought only a list of records returned from the PACER search, [since] the list was not previously created or obtained by the agency . . . an order that defendant produce such a list would be tantamount to requiring defendant to create an agency record-something that FOIA does not mandate.” *Id.*

Later, in *National Security Counselors v. CIA*, the same court extended the dicta in *American Way* to an internal agency database. 898 F. Supp. 2d at 271–72. The court admitted that the distinction between searching databases and creating new records was “somewhat muddled.” *Id.* at 270. It nevertheless attempted to draw a distinction between producing “particular points of data” on one hand and producing “listing or index” of a database on the other. *Id.* at 271. The former, the court held, was required under FOIA because the individual data points were “the records themselves,” but the latter would require the creation of a new record. *Id.* at 271. Confusingly, though, the example the court gave of a request that would require the creation of a new record was “a database listing of the first 100 FOIA requests filed in Fiscal Year 2012,” *id.*, a request that would appear to only require searching and sorting the information in the database. *See infra* Part II.C.

The content-index distinction outlined in *National Security Counselors* has found traction outside of the database context. *See, e.g., Johnson v. CIA*, 330 F. Supp. 3d 628, 642–43 (D. Mass. 2018); *Landmark Legal Found. v. U.S. Dep’t of Justice*,

211 F. Supp. 3d 311, 318 (D.D.C. 2016); *Tereshchuk v. Bureau of Prisons*, 67 F. Supp. 3d 441, 451–52 (D.D.C. 2014), *aff'd sub nom. Tereshchuk v. Bureau of Prisons, Dir.*, No. 14-5278, 2015 WL 4072055 (D.C. Cir. June 29, 2015). But state courts that have used *National Security Counselors* as persuasive authority have noted that the distinction between content and index is arbitrary and difficult to apply to databases in practice. *See, e.g., Am. Civil Liberties Union v. Arizona Dep't of Child Safety*, 240 Ariz. 142, 150 (Ct. App. 2016) (“We acknowledge that distinguishing between searching an electronic database and creating a new record that compiles previously un-compiled information about information may be a difficult task.”). The case currently before this Court represents an important opportunity to clarify this murky doctrinal area.

B. Accessing a particular arrangement of information by querying a database does not amount to creation of a new record.

The view that database queries in general involve accessing existing records, rather than creating new records, is consistent with FOIA jurisprudence. FOIA obligates an agency “to provide access to those [records] which it in fact has created and retained.” *Kissinger v. Reporters Comm. for Freedom of the Press*, 445 U.S. 136, 152 (1980); *Yeager*, 678 F.2d at 321. When an agency chooses to store information in a database, it is choosing to keep that information in a highly structured format that can be queried in many ways. That, in essence, is the entire point of keeping information in a database: the agency does not need to commit to a particular

arrangement of information. Rather, it has a multitude of different arrangements at its fingertips, each of which is in the agency's "possession or control." *Kissinger*, 445 U.S. at 152. To say that some of these available arrangements "have been *in fact* obtained" by the agency, *Forsham*, 445 U.S. at 186, but that others have not, would amount to arbitrary line-drawing.

Nor does this approach conflict with the rule that "FOIA imposes no duty on the agency to create records." *Id.* Creation of a new record occurs when an agency produces information that is not already affixed in an existing record. *See, e.g., Sears*, 421 U.S. at 161–62 (agency not required to draft explanatory material); *Tereshchuk*, 67 F. Supp. 3d at 451–52 (agency not required to add additional detail to existing records). On the other hand, a query that extracts a particular arrangement of information from data already affixed in a database does not constitute creation of a new record. The creation of a new record happens when data is added to a database. A query simply extracts that information in a usable, if potentially novel, form. *See Schladetsch*, 2000 WL 33372125, at *3 ("The fact that the agency may have to search numerous records to comply with the request and that the net result of complying with the request will be a document the agency did not previously possess is not unusual in FOIA cases, nor does this preclude the applicability of the Act.").

Treating database queries as accessing existing records, rather than creating new records, is also consistent with the Department of Justice's own guidance. In

discussing when it is appropriate to subdivide a document into constituent records, the Department provides the following example:

[I]f a document consists of a list of summaries of complaints against immigration judges organized by the name of each judge, and the subject of the FOIA request is “complaints against all immigration judges” then the entire document is the “record” for purposes of that FOIA request because the entire document is a “collection or grouping of information” on the subject of the request. Conversely, if the FOIA request specifically concerns “complaints against Judge Smith,” then only the complaint summaries concerning Judge Smith would constitute the “record” for purposes of that FOIA request, as only those portions would be the “collection or grouping of information” on Judge Smith.

Office of Information Policy, *Defining a “Record” Under the FOIA*, U.S. Dep’t of Justice (last updated Feb. 15, 2017) [hereinafter “*OIP Guidance*”].⁶ Nowhere in this example or the remaining guidance does the OIP suggest that subdividing a multi-subject document into discrete records constitutes the creation of new records.

The same reasoning applies to queries run on a database. A query extracts the portions of a database that constitute a “collection or grouping of information” matching the given criteria. A query may return a result set that happens to exactly match a grouping of data that is defined in the database schema, such as a table, row, or column. However, a query may also return a result set that matches a conceptual grouping of information that is not explicitly defined in the database schema. This

⁶ Available at https://www.justice.gov/oip/oip-guidance/defining_a_record_under_the_foia [<https://perma.cc/VA6G-JXMJ>].

could be the case when a database is normalized, and a query combines information that is spread across multiple tables. It could also be the case when a query returns aggregated data, such as the sum, average, or count of multiple rows or columns. In either case, the fact that a particular result set is not explicitly defined in the database schema does not mean that result set is a “new” record for the purposes of FOIA. Rather, any query that organizes existing information in a database into a “collection or grouping” should be treated as the retrieval of existing records.

Defining records in this way does not foreclose other potential interpretations of records. The OIP Guidance provides that “[t]he nature of a FOIA ‘record’ is defined by both the content of a document *and* the subject of the request.” *Id.* (emphasis in original). Applying this guidance to a database such as the FTS, it is apparent that in the context of some requests the entire database should be treated as a single record. This would be appropriate when the subject of the request is broad enough to encompass the subject matter of the database generally. *Id.* (noting that an “entire document is [a] ‘record’ for purposes of [a] FOIA request [when] the entire document is a ‘collection or grouping of information’ on the subject of the request”).

But databases are also good candidates for subdivision into constituent records when a request is more narrowly defined. As explained in Part I.A, *supra*, databases store information in a highly structured format that is easily divided and

recombined into a variety of arrangements. See *OIP Guidance* (explaining that subdivision of a record would be appropriate where record “readily lent itself to division according to [the] criterion [in the request]”). Thus, just as “an entire string of emails, a single email within a string of emails, or a paragraph within a single email could potentially constitute a ‘record’ for purposes of the FOIA,” *id.*, so could an entire database or particular subset of data within a database. Approaching queries as methods of retrieving records in the context of a request like CIR’s is most consistent with the Department of Justice’s own recommendation that agencies take “fine-tuned, content-based approach to the decision” of what constitutes a record, “irrespective of the physical attributes” of how the record is stored. *Id.*

C. Treating queries as accessing existing records avoids absurd results.

As described above, some courts have attempted to distinguish between queries that access existing records and queries that create new records. *See, e.g., Nat’l Sec. Counselors*, 898 F. Supp. 2d at 270–71. This amounts to holding that some arrangements of the information stored in a database already exist, but other arrangements of the same information are “new.” For the reasons stated above, this undertaking fundamentally misapprehends the nature of information stored in databases and how that information is accessed. Moreover, it inevitably leads to absurd results.

At the most basic level, a court might distinguish existing records from new records by only requiring agencies to turn over arrangements of information that have been extracted from a database and stored in some other format. However, this would effectively cut off access to all records stored in databases, rendering the E-FOIA Amendments moot. Such an approach would also conflict with the general rule that FOIA is to be construed broadly to provide access to records. *See U.S. Dep't of State v. Ray*, 502 U.S. 164, 173 (1991). Therefore, FOIA demands something more than release of records that have previously been extracted from a database.

Alternatively, a court might hold that FOIA requires agencies to release the record created by any query that had already been run on a database, but not to create new queries. This approach is problematic, though, because databases do not necessarily store a complete history of all queries that have even been run. Therefore, this solution would actually create *more* work for agencies, by requiring them to log each query that is run against a database.

Some courts have held that queries constitute the creation of new records if they involve “analysis,” “research,” or the generation of an “index.” *See Nat'l Sec. Counselors*, 898 F. Supp. 2d at 269–70; *People for Am. Way*, 451 F. Supp. 2d at 15; *Thompson Pub. Grp.*, 1994 WL 116141, at *2. For example, the court in *National Security Counselors* held that although “searching or sorting an electronic database is clearly not the creation of a record under the FOIA,” a query seeking “a database

listing of the first 100 FOIA requests filed in Fiscal Year 2012” would nevertheless amount to the creation of a new record. 898 F. Supp. 2d at 271.

In database terms, however, that request is nearly identical to a search. Consider a simple database that contains a table called “Requests” with a column called “Date” that recorded the date on which a request was filed. A query to find all of the requests for 2012 would look something like this:

```
SELECT * FROM Requests WHERE YEAR(Date)= 2012
```

This is a simple search, and would “clearly” not constitute the creation of a new record under the reasoning of *National Security Counselors*. The query the court found objectionable, though, would look almost identical:

```
SELECT * FROM Requests WHERE YEAR(Date)= 2012 ORDER BY Date LIMIT 100
```

The addition of a single clause at the end of the query would be sufficient to extract the requested arrangement of data. The same is true of aggregate data. The query to extract the number of requests per year would look something like this:

```
SELECT COUNT(*) FROM Requests GROUP BY YEAR(Date)
```

As these examples show, there is simply no meaningful difference in complexity between a search query and a query that produces an index or aggregate results.

Finally, technical distinctions between types of queries could needlessly complicate the FOIA process. For example, a court might rule that only queries that

return whole tables, rows, or columns are required under FOIA. But in the case of a normalized database, returning a single record may require querying across multiple tables. Would this be required under FOIA, or would joining multiple tables constitute creation of a new record? Technical rules like this would require FOIA requesters to understand in detail the inner workings of an agency database in order to make appropriate requests, and require courts to make factual determinations about the structure of agency databases to resolve FOIA disputes. Such an outcome would be detrimental to the entire FOIA process.

III. Treating queries as accessing existing records does not inappropriately expand the obligations of agencies under FOIA.

If a query constitutes accessing an existing record, then agencies are required to run queries against databases in their possession when conducting a search pursuant to a FOIA request. However, just as “[t]he type of storage system in which the agency has chosen to maintain its records cannot diminish the duties imposed by the FOIA,” *Yeager*, 678 F.2d at 321, “the fact that the requested information could be compiled by using information in an agency’s database should not expand an agency’s duty under the FOIA.” *Thompson Pub. Grp.*, 1994 WL 116141, at *1. Accordingly, *amici* note two important limitations on an agency’s duty to produce records by querying a database.

First, under the approach advanced by *amici*, agencies would not be required to run queries that incorporate information that is not already part of the database, as

doing so would in fact entail the creation of a new record. For example, assume that instead of requesting gun traces aggregated by year, the appellant here had requested the number of gun traces that involved violent crimes. If the FTS database included a field to indicate whether a particular crime was violent, such a request would have to be fulfilled, as doing so would amount to querying the database for an existing record. If, however, there was no such field in the database, the agency could not be compelled to add it, as doing so would entail the creation of a new record.

Second, agencies would not be required to run queries that create an unreasonable burden on the agency. It is a long-standing rule that “FOIA does not require an agency to conduct what amounts to an unreasonably burdensome search in response to a request.” *People for Am. Way*, 451 F. Supp. 2d at 12. This rule applies in the database context as well. Of course, one of the purposes of storing information in a database is that highly precise queries can be carried out with minimal effort. *See Nat’l Sec. Counselors*, 898 F. Supp. 2d at 272 (noting that querying a database may result in overall lower burden on the agency than producing the entire database). Unreasonably burdensome queries are therefore unlikely. However, it remains the case that if an agency can demonstrate that a particular database query would create an unreasonable burden, due to the complexity of the query or some other factor, then the agency need not run that query.

In summary, rather than make artificial distinctions about which queries are required under FOIA and which are not, the Court should hold that any arrangement of information that can be retrieved from a database with a simple query or series of queries constitutes an existing record for FOIA purposes.

CONCLUSION

For the reasons stated above, *amici* respectfully urge this Court to REVERSE the decision below with regard to the creation of new records, hold that retrieving a record from a database does not constitute the creation of a new record, and remand for further proceedings consistent with this holding.

Respectfully submitted,

Dated: March 28, 2019

s/ Mason A. Kortz

Mason A. Kortz
Kendra K. Albert
Cyberlaw Clinic⁷
Harvard Law School
1585 Massachusetts Ave.
Cambridge, MA 02139
(617) 495-2845
mkortz@law.harvard.edu

Counsel for *amici curiae*

⁷ *Amici curiae* thank Spring 2019 Cyberlaw Clinic students Alena Farber, Ariel Hoffman, and LeHeng Li for their valuable contributions to this brief. *Amici* also thank Ted Han for his support in bringing together this coalition of media organizations and data journalists and Evan Gallo for his Harry Potter knowledge.

CERTIFICATE OF COMPLIANCE

Pursuant to the Federal Rule of Appellate Procedure 32(a)(7)(C), I hereby certify that:

This brief complies with the type volume limitations of Federal Rules of Appellate Procedure 29(a)(5) and 32(a)(7)(b) and Ninth Circuit Rule 32-1(a) because it contains 6,958 words as calculated by the word count feature of Microsoft Office 365, exclusive of the sections exempted by Federal Rule of Appellate Procedure 32(f).

This brief complies with the typeface requirement of Federal Rule of Appellate Procedure 32(a)(5)(A) and (a)(6) because it uses 14-point proportionally spaced Times New Roman font.

Dated: March 28, 2019

s/ Mason A. Kortz

Mason A. Kortz
Cyberlaw Clinic
Harvard Law School
1585 Massachusetts Ave.
Cambridge, MA 02139
(617) 495-2845
mkortz@law.harvard.edu

CERTIFICATE OF SERVICE

I hereby certify that I electronically filed the foregoing Brief of *Amici Curiae* Five Media Organizations and Sixteen Data Journalists in Support of Appellant and Reversal with the Clerk of the Court for the United States Court of Appeals for the Ninth Circuit by using the appellate CM/ECF system on March 28, 2019. I certify that all participants in the case are registered CM/ECF users and that service will be accomplished by the appellate CM/ECF system.

Dated: March 28, 2019

s/ Mason A. Kortz

Mason A. Kortz
Cyberlaw Clinic
Harvard Law School
1585 Massachusetts Ave.
Cambridge, MA 02139
(617) 495-2845
mkortz@law.harvard.edu