

COMMONWEALTH OF MASSACHUSETTS  
**SUPREME JUDICIAL COURT**

---

No. SJC-12622

---

BOSTON GLOBE MEDIA PARTNERS, LLC  
Plaintiff/Appellant

v.

DEPARTMENT OF PUBLIC HEALTH  
Defendant/Appellee

---

On Appeal from a Judgment  
of the Suffolk County Superior Court

---

**BRIEF OF THE REPORTERS COMMITTEE FOR FREEDOM OF THE  
PRESS, THE EDITORIAL STAFF OF *THE TECH*, METRO CORP.,  
THE NEW ENGLAND CENTER FOR INVESTIGATIVE REPORTING,  
THE NEW ENGLAND FIRST AMENDMENT COALITION, THE NEW  
ENGLAND NEWSPAPER AND PRESS ASSOCIATION, THE NEW YORK  
TIMES COMPANY, NORTH OF BOSTON MEDIA GROUP, AND THE  
EDITORIAL STAFF OF *THE FREE PRESS* AS *AMICI CURIAE* IN  
SUPPORT OF PLAINTIFF/APPELLANT**

---

Andrew F. Sellars (BBO #682690)  
**BU/MIT TECHNOLOGY & CYBERLAW CLINIC**  
Boston University School of Law  
765 Commonwealth Avenue  
Boston, MA 02215  
Tel: (617) 358-7377  
Fax: (617) 353-6944  
sellars@bu.edu

Date: December 21, 2018



On the brief:

Julissa Milligan  
BU/MIT Technology & Cyberlaw Clinic  
Boston University School of Law  
765 Commonwealth Ave.  
Boston, MA 02215  
Tel: (617) 353-3140  
jlmillig@bu.edu

Bruce D. Brown  
Katie Townsend  
Caitlin Vogus  
Reporters Committee for Freedom of the Press  
1156 15th St. NW, Ste. 1020  
Washington, DC 20005  
Tel: (202) 795-9300  
bbrown@rcfp.org



**TABLE OF CONTENTS**

Table of Authorities ..... ii  
Statement of Issues Addressed by *Amici Curiae* ..... 1  
Statement of Interest of *Amici Curiae* ..... 1  
Statement of the Case and Relevant Facts ..... 5  
Summary of Argument ..... 7  
Argument ..... 10

I. THE DEPARTMENT FAILED TO DEMONSTRATE WITH PARTICULARITY THAT RELEASING THE RECORDS MAY CAUSE AN UNWARRANTED INVASION OF PERSONAL PRIVACY. .... 10

A. The Department failed to show that releasing an electronic compilation of Records that are already available for public inspection will cause harm. .... 11

B. The Department’s authorities provide no support for its contention that disclosure of the Records will cause serious privacy harms. .... 14

II. THE SUPERIOR COURT IMPROPERLY SUBSTITUTED ITS OWN JUDGMENT FOR THAT OF THE LEGISLATURE WHEN IT PROHIBITED THE DISCLOSURE OF RECORDS THAT THE LEGISLATURE DETERMINED SHOULD BE PUBLIC. .... 18

III. THE COURT’S EXISTING FRAMEWORK FOR ANALYZING PRIVACY CLAIMS IS SUFFICIENTLY ROBUST TO HANDLE THE CHALLENGES POSED BY “BIG DATA.” ..... 20

A. Courts and custodians should distinguish between the “breadth” and the “depth” of the record set in question. .... 22

B. These records do not pose a privacy concern, despite being “broad.” ..... 24

IV. RELEASE OF STATEWIDE VITAL STATISTICS RECORDS HAS NUMEROUS PUBLIC BENEFITS. .... 26

Conclusion ..... 32

**TABLE OF AUTHORITIES**

**CASES**

*Att’y Gen. v. Assistant Comm’r of Real Property Dep’t*,  
380 Mass. 623 (1980).....13

*Boston Globe Media Partners, LLC v. Dep’t of Pub.  
Health, Suffolk Cty. Sup. Ct., No. 2014-4074-E  
(Aug. 25, 2017)*.....7, 12

*Cape Cod Times v. Sheriff of Barnstable Cty.*,  
443 Mass. 587 (2005).....8, 11, 13, 23

*Champa v. Weston Pub. Schs.*, 473 Mass. 86 (2015)  
.....8, 11, 12, 22

*Clymer v. City of Cedar Rapids*, 601 N.W.2d 42  
(Iowa 1999).....16

*Data Tree, LLC v. Meek*, 109 P.3d 1226 (Kan. 2005)  
.....15-16

*Dep’t of State v. Wash. Post Co.*,  
456 U.S. 595 (1982).....14

*Georgiou v. Comm’r of Dep’t of Indus. Accidents*,  
67 Mass. App. Ct. 428 (2006).....22

*Globe Newspaper Co. v. Boston Ret. Bd.*,  
388 Mass. 427 (1983).....22, 26

*Globe Newspaper Co. v. Police Comm’r of Boston*,  
419 Mass. 852 (1995).....12

*In re Colokathis*, 417 B.R. 150 (Bankr. D. Mass. 2009)  
.....15

*Pottle v. Sch. Comm. of Braintree*,  
395 Mass. 861 (1985).....13, 23

*Suffolk Const. Co. v. Div. Of Capital Asset Mgt.*,  
449 Mass. 444 (2007).....11

*Texas Comptroller of Public Accounts v. Att’y General  
of Texas*, 354 S.W.3d 336 (Tex. 2010) .....16

*True the Vote v. Hosemann*,  
43 F. Supp. 3d 693 (S.D. Miss. 2014).....16

*Wakefield Teachers Ass’n v. Sch. Comm. of Wakefield*,  
431 Mass. 792 (2000).....19

*Worcester Telegram & Gazette Corp. v. Chief of Police of Worcester*, 436 Mass. 378 (2002).....8, 11

*Worcester Telegram & Gazette Corp. v. Chief of Police of Worcester*, 58 Mass. App. Ct. 1 (2003).....19

**STATUTES AND LEGISLATIVE MATERIAL**

1887 Mass. Laws, c. 202 §§ 1-2 ..... 20

2015 House Doc. No. 4333 ..... 11

G.L. c. 4, § 7, cl. 26 ..... *passim*

G.L. c. 46, § 2 ..... 20

G.L. c. 46, § 2A ..... 20

G.L. c. 46, § 2B ..... 20

G.L. c. 46, § 2C ..... 20

G.L. c. 46, § 13 ..... 20

G.L. c. 66, § 10 (2010) ..... 11

G.L. c. 66, § 10A ..... 11

5 U.S.C. § 552 ..... 13

**OTHER AUTHORITIES**

Alessandro Acquisti & Ralph Gross, *Predicting Social Security Numbers From Public Data*, 106 Proc. Nat’l Acad. Sci. 10975 (2009), available at <https://www.pnas.org/content/106/27/10975>..... 17

Alessandro Acquisti & Ralph Gross, *SSN Study - FAQ*, Carnegie Mellon University (July 29, 2009), <https://www.heinz.cmu.edu/~acquisti/ssnstudy/>..... 17

Micah Altman et al., *Towards a Modern Approach to Privacy-Aware Government Data Releases*, 30 Berkley Tech. L.J. 1967(2016)..... 23

D. Victoria Baranetsky, *Data Journalism and the Law* (Sept. 19, 2018), Tow Ctr. for Digital Journalism, [https://www.cjr.org/tow\\_center\\_reports/data-journalism-and-the-law.php/](https://www.cjr.org/tow_center_reports/data-journalism-and-the-law.php/)..... 27

Michael Barthel, *5 Facts About the State of the News Media in 2017*, Pew (Aug. 21, 2018), <http://www.pewresearch.org/fact-tank/2018/08/21/5-facts-about-the-state-of-the-news-media-in-2017/> ..... 32

Sec. William Francis Galvin, Div. of Pub. Records, *A Guide to Massachusetts Public Records Law* (2017) ..... 15

Anne Glover & David Beard, *Study Shows Data Reporting Gaining Hold in Newsrooms*, Poynter (Sept. 18, 2017), <https://www.poynter.org/tech-tools/2017/study-shows-data-reporting-gaining-hold-in-newsrooms/>..... 28

David Goldstein, *CBS2 Investigation Uncovers Votes Being Cast From Grave Year After Year*, CBS Los Angeles (May 23, 2016), <https://losangeles.cbslocal.com/2016/05/23/cbs2-investigation-uncovers-votes-being-cast-from-grave-year-after-year/>..... 28

William Heisel, *A Public Death: Digital Records Could Combat Identity Theft*, Ctr. for Health Journalism (Feb. 8, 2012), <https://www.centerforhealthjournalism.org/blogs/2012/02/08/public-death-digital-records-could-combat-identity-theft>..... 18

William Heisel, *Well Sourced: Use Vital Records to Find Patients, Document Trends*, Ctr. for Health Journalism (May 5, 2015), <https://www.centerforhealthjournalism.org/2015/05/04/well-sourced-use-vital-records-find-patients-document-trends>..... 30

Daniel T. Larose & Chantal D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining* (2d ed. 2014)..... 24, 25

Boris Lubarsky, *Re-Identification of "Anonymized" Data*, 1 Geo. Tech. L. Rev. 202 (2017)..... 25

Adanya Lustig, *In Rural Hospitals, Women are Dying During Childbirth – and States Aren't Keeping Track*, MuckRock (Dec. 4, 2017), <https://www.muckrock.com/news/archives/2017/dec/04/maternal-deaths-map/>..... 29

John Markoff, *Weakness in Social Security Numbers Found*, N.Y. Times, July 6, 2009, available at <https://www.nytimes.com/2009/07/07/us/07numbers.html> ..... 18

Mark Memmott, *'Memmos': Memmott's Missives & Musings*, NPR Ethics Handbook, (Nov. 20, 2014), <http://ethics.npr.org/category/memos-from-memmott/#1984>..... 27

*Untold Stories: A Survey of Freelance Investigative Reporters*, ProjectWorld (2015), [http://www.firenewsroom.org/sites/default/files/PW\\_Freelancer\\_Survey\\_2015.pdf](http://www.firenewsroom.org/sites/default/files/PW_Freelancer_Survey_2015.pdf)..... 32

*All Data Sets*, ProPublica, <https://www.propublica.org/datastore/datasets>... 27-28

*FOI A to Z*, Society of Professional Journalists, <https://www.spj.org/opendoors7.asp>..... 30

Emily Sweeney, *A New Window on Bay State's Vital Records*, Boston.com (March 20, 2012), [http://archive.boston.com/news/local/massachusetts/articles/2012/03/20/vital\\_records\\_go\\_from\\_machusetts\\_town\\_vaults\\_to\\_cyberspace/](http://archive.boston.com/news/local/massachusetts/articles/2012/03/20/vital_records_go_from_machusetts_town_vaults_to_cyberspace/)..... 20

Anjali Tsui, *Delaware Becomes First State to Ban Child Marriage*, Frontline (May 9, 2018), <https://www.pbs.org/wgbh/frontline/article/delaware-becomes-first-state-to-ban-child-marriage/>..... 29

Giske Ursin et al., *Protecting Privacy in Large Datasets - First We Assess the Risk; Then We Fuzzy the Data*, 26 Cancer Epidemiology, Biomarkers and Prevention 1219(2017)..... 25-26

Steph Yin, *When Did Americans Stop Marrying Their Cousins? Ask the World's Largest Family Tree*, N.Y. Times (March 1, 2018), <https://www.nytimes.com/2018/03/01/science/cousins-marriage-family-tree.html>..... 30

### **Statement of Issues Addressed by *Amici Curiae***

The Public Records Law ("PRL") favors disclosure of governmental records, but excludes disclosure of "materials or data relating to a specifically named individual, the disclosure of which may constitute an unwarranted invasion of personal privacy[.]" G.L. c. 4, § 7, cl. 26(c) ("Exemption (c)"). This Court has solicited *amicus curiae* briefs to address whether the Superior Court was correct to apply this exemption to withhold an electronic index of birth and marriage records.

### **Statement of Interest of *Amici Curiae***

The Reporters Committee for Freedom of the Press is an unincorporated nonprofit 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.

Emma Bingham is the Editor in Chief, and Fiona Chen, Charlie Moore, Áron Ricardo Perez-Lopez, Nafissa

Syed, Steven Truong, Patrick Wahl, and Vivian Zhong are members of the editorial staff of *The Tech*, the oldest and largest student newspaper at the Massachusetts Institute of Technology ("MIT").<sup>1</sup> *The Tech* has been in continuous operation since 1881, and produces original journalism throughout the year on issues of concern to MIT students, faculty, and staff, as well as citizens in Cambridge and the greater Boston area. As a long-standing small news organization, *The Tech* has a vital interest in ensuring that the traditional methods of newsgathering through public records requests are made meaningful as technology evolves and information is presented in new formats and media.

Metro Corp., the publisher of *Boston* magazine, is the nation's second largest publisher of city magazines. *Boston* magazine is published monthly and has been reporting on Boston's cultural and political trends since 1963. Metro Corp. is a privately held corporation owned primarily by David H. Lipson and trusts that were established for the benefit of his

---

<sup>1</sup> The individuals appear in their personal capacity. *The Tech* is a student organization within the Massachusetts Institute of Technology, who is not a party to this brief.

heirs and the heirs of the late D. Herbert Lipson. No publicly held corporation owns 10% or more of Metro Corp.'s stock.

The New England First Amendment Coalition ("NEFAC") is a nonpartisan, nonprofit organization that advances and protects the five freedoms of the First Amendment, including the public's right to know about its government. It is the region's leading press freedom and open government advocate, serving Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island and Vermont. In addition to defending the rights of New Englanders, NEFAC educates those across the country and the world about the First Amendment, the value of transparency and the importance of a free press. The coalition regularly files and joins *amicus curiae* briefs in cases involving the First Amendment and the public's right to know. It has no parent company and no stock.

The New England Center for Investigative Reporting ("NECIR") is a nonprofit journalism center based at Boston University and with offices at the studios of WGBH News in Boston, Massachusetts. NECIR produces high-impact, public interest-oriented investigative reporting written by experienced,

professional journalists with the assistance of student researchers. NECIR's stories appear in newspapers and on television and radio stations across Massachusetts, New England, and the nation. NECIR is wholly owned by the Trustees of Boston University, a nonprofit educational corporation.

The New England Newspaper and Press Association, Inc. ("NENPA") is the regional association for newspapers in the six New England States (including 210 in Massachusetts). NENPA's corporate office is in Woburn, Massachusetts. Its purpose is to promote the common interests of newspapers published in New England. Consistent with its purposes, NENPA is committed to preserving and ensuring the open and free publication of news and events in an open society. New England Newspaper and Press Association, Inc. is a non-profit corporation. It has no parent, and no publicly held corporation owns 10% or more of its stock.

The New York Times Company is the publisher of *The New York Times* and *The International Times*, and operates the news website [nytimes.com](http://nytimes.com). The New York Times Company is a publicly traded company and has no

affiliates or subsidiaries that are publicly owned. No publicly held company owns 10% or more of its stock.

North of Boston Media Group is the publisher of eight daily and weekly newspapers in northeastern Massachusetts and southern New Hampshire, including the *Eagle-Tribune*, which has been in operation since 1868. North of Boston Media Group is a wholly-owned subsidiary of CNHI, LLC.

Julie Pike is the Editor in Chief, and Dionne Smith, Lydia Libby, Asha Tompkins, and Laura McCallum are members of the editorial staff of *The Free Press*, the official student newspaper of the University of Southern Maine.<sup>2</sup> In reporting on both campus news and issues relating to the greater Portland and Lewiston-Auburn communities, the staffers of *The Free Press* are keenly aware of the importance of access to public records. *The Free Press* thus has an interest in the ability to access these records in New England states.

#### **Statement of the Case and Relevant Facts**

In May 2013, Plaintiff-Appellant Boston Globe Media Partners, LLC (the "Globe") made a PRL request

---

<sup>2</sup> The individuals appear in their personal capacity. *The Free Press* is a student organization within the University of Southern Maine, who is not a party to this brief.

to the Registry of Vital Records and Statistics ("Registry"), within the Defendant-Appellee Department of Public Health (the "Department"). The request sought, *inter alia*, an electronic copy of indices of births and marriages (collectively, the "Records"). Globe Br. 13. The birth index lists the subject's first and last name, date of birth, place of birth, names of parents, and location of the full record in the Registry vault. *Id.* at 7. The marriage index lists both spouses' first and last name, the date of marriage, the place where the license was filed, the marriage certificate number, and the location of the full record in the Registry vault. *Id.* at 7-8.

All of the information found in the Records can be retrieved by searching state-owned public computer terminals in the Registry's research room, but it is not available in aggregate form. *Id.* at 4-5. Using these terminals costs \$9.00 per hour. *Id.* at 5. It would take countless hours, and be prohibitively expensive, to retrieve all of the Records using the terminals.

The Department denied the Globe's request, arguing that the Records are exempt from disclosure under Exemption (c) and Exemption (a), the latter of

which allows a department to withhold records “specifically or by necessary implication exempted from disclosure by statute,” G.L. c. 4, § 7, cl. 26(a) (“Exemption (a)”).<sup>3</sup> The Globe appealed to the Supervisor of Public Records, who twice stated that the Records should be released, before reversing that position after the Department requested reconsideration. Globe Br. 14-15.

The Globe filed the present action, seeking a declaratory judgment that the Records are not exempt from disclosure under the PRL. *Id.* at 1-2. After argument on cross-motions for summary judgment, the Superior Court granted the Department’s motion. The court found the Records are not exempt under Exemption (a), but are exempt under Exemption (c). *Boston Globe Media Partners, LLC v. Dep’t of Pub. Health*, Suffolk Cty. Sup. Ct., No. 2014-4074-E (Aug. 25, 2017). The Globe then filed the present appeal. Globe Br. 2.

### **Summary of the Argument**

The letter and spirit of the PRL is to “give the public broad access to governmental records.”

---

<sup>3</sup> *Amici* agree with Plaintiff-Appellant’s position that the Records are not subject to withholding under Exemption (a), but write specifically to address the application of Exemption (c).

*Worcester Telegram & Gazette Corp. v. Chief of Police of Worcester*, 436 Mass. 378, 382-383 (2002). The Department reads this precisely backwards. It argues that because the records are broad, the public should not be given access. See Dep't Br. 39. The Department does not dispute that the Records are rightly available to the public at an individual level. *Id.* at 21. Aggregating non-intimate, publicly-available individual records does not inherently transform the compilation into highly intimate or embarrassing information.

The government failed to meet its burden under the PRL, and misapplied both judicial decisions and data science. The Department cannot and does not show that the Records include highly intimate personal information, or that their disclosure will embarrass a reasonable person. *Champa v. Weston Pub. Schs.*, 473 Mass. 86, 96-97 (2015); *Cape Cod Times v. Sheriff of Barnstable Cty.*, 443 Mass. 587, 594-595 (2005). Nor do the Department's legal authorities substantiate its position. Some concerned far more invasive compilations of personal data, or simply failed to substantiate the alleged harm. In others, courts

misconstrued academic studies that examined this question. See Section I.B, *infra*.

Moreover, the Department and the Superior Court's decision below disregarded the legislature's unequivocal policy judgment – reached more than 130 years ago – that these types of records should be available for public inspection. See Section II, *infra*.

This Court should not abandon its well-settled framework for evaluating exemption requests predicated on privacy concerns under the PRL simply because the Records contain numerous entries. To the extent that large datasets are different than other records, they can be analyzed simply by distinguishing between the “breadth” and the “depth” of the dataset in question. “Broad” but “shallow” datasets, like the Records here, which relate to numerous individuals but contain few details, pose much lower privacy risks than “deep” datasets of any breadth, which contain detailed information about each person in the set. See Section III.A, *infra*. In this case, because these Records contain very little information about each person, disclosing the Records will not create any new privacy risks. See Section III.B, *infra*.

Finally, the public benefit in releasing the Records is considerable. This dataset can be used to facilitate government accountability, gain new insights into systemic questions about the welfare of Massachusetts citizens, and inform localized news reporting by news outlets of all sizes. These benefits are not possible if the Department is permitted to paywall access to these Records and refuse to produce the full set under the PRL. See Section IV, *infra*.

For these reasons, *amici* respectfully request that this Court reverse the Superior Court, and order inspection and disclosure of the Records.

### **Argument**

#### **I. THE DEPARTMENT FAILED TO DEMONSTRATE WITH PARTICULARITY THAT RELEASING THE RECORDS MAY CAUSE AN UNWARRANTED INVASION OF PERSONAL PRIVACY.**

Electronic recordkeeping has not changed the fundamental fact that the Records the Globe seeks are subject to public disclosure. There is no “big data” exemption in the PRL.

The Department relies on unsubstantiated, conclusory assertions to suggest that disclosing the Records will cause an unwarranted invasion of personal privacy, see Dep’t Br. 46-48, but these arguments do not satisfy the Department’s burden to demonstrate

that the Records will expose highly intimate personal information, or that their disclosure will embarrass a reasonable person. *Champa*, 473 Mass. at 96-97; *Cape Cod Times*, 443 Mass. at 594-595. Even the cases the Department relies on fail to support its claims.

**A. The Department failed to show that releasing an electronic compilation of records that are already available for public inspection will cause harm.**

The "primary purpose" of the PRL, firmly rooted in case law and statute, "is to give the public broad access to public records." *Worcester Telegram & Gazette Corp.*, 436 Mass. at 382-383. Records are presumptively public, and the government bears the burden to "prove with specificity" the statutory grounds for withholding. G.L. c. 66, § 10(c) (2010); see also *Suffolk Const. Co. v. Div. Of Capital Asset Mgt.*, 449 Mass. 444, 454 (2007).<sup>4</sup>

---

<sup>4</sup> The parties agree that this case is governed by the PRL as it existed before its most recent amendment, effective in 2017. In that amendment, the General Court changed the statute from requiring the government to demonstrate that an exemption applies with "particularity," to requiring proof by a "preponderance of the evidence." G.L. c. 66, § 10A(d)(1)(iv) (2017). There is no evidence that the legislature intended to change the substantive burden required through this restyling. See 2015 House Doc. No. 4333 (stating revision is "intended to bring . . . more specificity to the existing law").

To determine whether Exemption (c), which exempts from disclosure records that would invade a person's privacy, applies, Massachusetts courts apply a two-part test. First, the court must examine whether the information implicates a recognized privacy interest, considering specifically whether the records "contain[] intimate details of a highly personal nature," *Globe Newspaper Co. v. Police Comm'r of Boston*, 419 Mass. 852, 858 (1995) (internal quotations omitted), or that "the disclosure would result in personal embarrassment to an individual of normal sensibilities." *Champa*, 473 Mass. at 96. This is counterbalanced by "whether the same information is available from other sources." *Id.* Only after this threshold is met does a court consider whether the public interest in obtaining the information substantially outweighs the seriousness of the invasion of privacy. *Id.* (citing *Att'y Gen. v. Collector of Lynn*, 377 Mass. 151, 156 (1979)).

In applying this test below, the Superior Court skipped directly to the balancing test and improperly found that the private harm outweighed the public interest in disclosure. *Boston Globe Media Partners, LLC*, No. 2014-4074-E, at 10-11. But the court should

not have even reached the balancing test here, because the Records do not meet the threshold test. See *Pottle v. Sch. Comm. of Braintree*, 395 Mass. 861, 865 (1985) (holding that the public employee home addresses do not fall under Exemption (c), without weighing the public interest in disclosure); *Cape Cod Times*, 443 Mass. at 594-595 (same).

Disclosing the Records will not implicate the privacy interests identified above. First, the parties agree that the government already makes the Records publicly available through individual searches, *Globe Br.* at 4-8; *Dep't Br.* at 21, so disclosure cannot cause an "unwarranted" invasion of privacy. Moreover, courts have consistently held that names and addresses are not private. See, e.g., *Pottle*, 395 Mass. at 865. While this Court previously analogized to the privacy provision in the federal Freedom of Information Act, 5 U.S.C. § 552, to suggest that marital status could be deemed sufficiently sensitive, see *Att'y Gen. v. Assistant Comm'r of Real Property Dep't*, 380 Mass. 623, 626 n.2 (1980) (citing *Rural Housing Alliance v. Dep't of Agric.*, 498 F.2d 73, 77 (D.C. Cir. 1974)), this Court has not directly addressed the question, and subsequent federal cases have contradicted this

earlier language. *Dep't of State v. Wash. Post Co.*, 456 U.S. 595, 602 n.5 (1982) (noting that the public nature of records that include "marital status" and "date of birth" "may be a reason to conclude . . . that the release of such information would not constitute a 'clearly unwarranted invasion of personal privacy'").

**B. The Department's authorities provide no support for its contention that disclosure of the Records will cause serious privacy harms.**

The Department suggests that the Records should not be disclosed because "data-linking and aggregation capabilities" will allow a malicious actor to use these Records for identity theft. The Department suggests that the Records should not be disclosed because, if combined with other, unidentified records, they could allow a third party to commit identity theft. *Dep't Br. 49*. However, the Department offers no support for its contention that the Records could be used to facilitate identity theft, and the cases it cites are inapposite.

The key datum of concern in an identity theft case is a social security number, which is not present

in the Records.<sup>5</sup> See *In re Colokathis*, 417 B.R. 150, 157 n.6 (Bankr. D. Mass. 2009) (identity theft tends to be either “new account fraud,” where a criminal opens up a line of credit using a real social security number paired with a real name, or “synthetic identity theft,” where a criminal uses a real social security number with a fake name) (citing Chris Jay Hoofnagle, *Identity Theft: Making the Unknown Known*, 21 Harv. J. L. & Tech. 97 (2007)). And none of the Department’s cases suggest that these Records could lead an identity thief to an individual’s social security number.

The Department cites a variety of cases from other jurisdictions to argue that the dissemination of birth dates and places of birth could be combined with information from other databases to facilitate identity theft. Dep’t Br. 43-49. But many of the cases it cites involved considerably more invasive datasets, including at least one that included social security numbers directly. See, e.g., *Data Tree, LLC v. Meek*,

---

<sup>5</sup> Nor does it seem likely that a social security number would be disclosed pursuant to a public records request. See William Francis Galvin, Div. of Pub. Records, *A Guide to Massachusetts Public Records Law* 25 (2017) (noting “social security numbers . . . may be withheld pursuant to Exemption (c)”).

109 P.3d 1226, 1238 (Kan. 2005) (data included birth dates, social security number, and mother's maiden names); *Clymer v. City of Cedar Rapids*, 601 N.W.2d 42, 48 (Iowa 1999) (data included gender, address, date of birth, and details regarding employee sick leave).

Some cases cited by the Department simply relied on unsupported assertions about identity theft. For example, one court simply asserted that "disclos[ing] . . . personal information such as birthdates" fuels identity theft "in the age of big data," without citing any source to support that assertion. *True the Vote v. Hosemann*, 43 F. Supp. 3d 693, 738 (S.D. Miss. 2014).

In other cases cited by the Department, courts misapplied research on the security of social security numbers. In *Texas Comptroller of Public Accounts v. Att'y General of Texas*, for example, the Texas Supreme Court relied on a *Dallas Morning News* article for the proposition that birth dates, "when combined with name and place of birth, can reveal social security numbers." 354 S.W.3d 336, 345 (Tex. 2010) (citing Bob Moos, *How Secure is Your Social Security Data?*, *Dallas Morning News*, Aug. 9, 2009, at 1D).

The article and the study it cites – which was also cited directly by the Department here, Dep’t Br. 47, n.18 – made a far more qualified claim. In the cited study from Carnegie Mellon University, researchers were able to determine *the first five digits* of a person’s social security number for 44% of the studied set, and could guess a social security number *in less than 1000* attempts for 8.5% of the studied records, with a higher probability of success for a few smaller states in some specific years (such as Delaware for those born in 1996). Alessandro Acquisti & Ralph Gross, *Predicting Social Security Numbers From Public Data*, 106 Proc. Nat’l Acad. Sci. 10975, 10975, 10978 (2009), available at <https://www.pnas.org/content/106/27/10975>. In a “frequently asked questions” document that accompanied the study, the authors rejected the suggestion that this increased predictability would lead to more identity theft, absent other exploitations. Alessandro Acquisti & Ralph Gross, *SSN Study - FAQ*, Carnegie Mellon University (July 29, 2009), <https://www.heinz.cmu.edu/~acquisti/ssnstudy/>.<sup>6</sup> And

---

<sup>6</sup> The Social Security Administration similarly appeared unperturbed by the study, with a spokesperson

indeed, others have asserted that the release of authentic digital records could actually help combat identity theft by dispelling false information. William Heisel, *A Public Death: Digital Records Could Combat Identity Theft*, Ctr. for Health Journalism (Feb. 8, 2012), <https://www.centerforhealthjournalism.org/blogs/2012/02/08/public-death-digital-records-could-combat-identity-theft> (widespread disclosure of death records helps prevent identity theft by providing a means of detecting the false use of another's identity).

The PRL requires a government to show with specificity that a sufficient privacy harm will flow from disclosure. The Department has not done so here.

**II. THE SUPERIOR COURT IMPROPERLY SUBSTITUTED ITS OWN JUDGMENT FOR THAT OF THE LEGISLATURE WHEN IT PROHIBITED THE DISCLOSURE OF RECORDS THAT THE LEGISLATURE DETERMINED SHOULD BE PUBLIC.**

The Superior Court correctly found that the legislature did not exempt the Records from disclosure

---

indicating that "[t]he method by which Social Security assigns numbers has been a matter of public record for years. The suggestion that [the study] has cracked a code for predicting an S.S.N. is a dramatic exaggeration." John Markoff, *Weakness in Social Security Numbers Found*, N.Y. Times, July 6, 2009, available at <https://www.nytimes.com/2009/07/07/us/07numbers.html>.

under Exemption (a). See G.L. c. 4, § 7, cl. 26(a). But this only tells half the story. The legislature has affirmatively decided that this specific information should be made public. By ignoring that fact, the Superior Court improperly substituted its own judgment for that of the legislature.

It is the legislature's mandate to "balance[] competing public policy considerations" and determine which records ought to be disclosed under public records laws; courts "shall not second-guess" the legislature's determination. *Wakefield Teachers Ass'n v. Sch. Comm. of Wakefield*, 431 Mass. 792, 802 (2000); see *Worcester Telegram & Gazette Corp. v. Chief of Police of Worcester*, 58 Mass. App. Ct. 1, 5 (2003) ("However cogent the parties' arguments might be as a matter of competing policy choices, the Legislature has [already] determined what records should be public and what exempt from disclosure.").

The legislature determined long ago that these Records should be publicly available. In 1887, the General Court decided that "any records of births, deaths and marriages" were required to be "preserved by [the town] clerk or registrar, and filed, arranged and indexed conveniently for examination and

reference." 1887 Mass. Laws, c. 202 §§ 1-2; see G.L. c. 46, § 2 (the current and highly similar language, as amended in 1968). The legislature has only carved out narrow departures from this general public right of access over the years, none of which are applicable here. See, e.g., G.L. c. 46, §§ 2A, 2B, 2C, 13. If the legislature wanted to prohibit the bulk disclosure of such records, it had ample opportunity and means to do so. It has not done so, even though bulk collection efforts have been underway for years.<sup>7</sup> There is no cause for courts to substitute their own policy preferences over the legislature's consistent belief that these records should be public.

**III. THE COURT'S EXISTING FRAMEWORK FOR ANALYZING PRIVACY CLAIMS IS SUFFICIENTLY ROBUST TO HANDLE THE CHALLENGES POSED BY "BIG DATA."**

The Court need not depart from its traditional framework for evaluating privacy-based exemption

---

<sup>7</sup> A 2012 article from the Globe told a story of an Oxford, Mass. couple who spent 30 years collecting bulk birth and death records from "almost every city hall and town hall in the state," before sending them to the website Ancestry.com for digitization, seemingly without significant privacy incident or legislative intervention. Emily Sweeney, *A New Window on Bay State's Vital Records*, Boston.com (March 20, 2012), [http://archive.boston.com/news/local/massachusetts/articles/2012/03/20/vital\\_records\\_go\\_from\\_machusetts\\_town\\_vaults\\_to\\_cyberspace/](http://archive.boston.com/news/local/massachusetts/articles/2012/03/20/vital_records_go_from_machusetts_town_vaults_to_cyberspace/).

claims under the PRL simply because this request seeks electronic records. The Department repeatedly references the "millions of private citizens" whose records will be disclosed, as if the fact that the Records contain numerous entries itself demonstrates a privacy harm. Record sets of this size may require courts and record custodians to take an additional moment to consider the impact of their disclosure, but they should not, and cannot, abandon existing privacy frameworks. Courts and custodians should instead consider (1) the "breadth" and the "depth" of the dataset in question; (2) whether the dataset is broad enough to present some unique additional considerations, as described below; and (3) whether having a similar dataset generally available elsewhere has caused harm.

This brief addresses the first two considerations at some length because the third point – that other governments and organizations routinely make records like these available in bulk and the Department has not shown that such disclosures caused any harm – has already been briefed by the parties. See Globe Br. 9-11.

**A. Courts and custodians should distinguish between the "breadth" and the "depth" of the record set in question.**

When assessing the privacy risks of disclosure of a particular dataset, a court should distinguish datasets that contain a "breadth" of entries (i.e., records about a large number of people) from datasets with a "depth" of entries (i.e., detailed or sensitive information about each person in the set). To date, this Court has largely considered "deep" record sets of varying breadth. *See Champa*, 473 Mass. at 87-88 (records of a single school district that revealed extensive information about students and their learning disabilities); *Globe Newspaper Co. v. Boston Ret. Bd.*, 388 Mass. 427, 429-430 (1983) (records of retired city employees that contained information about medical conditions and disabilities); *see also Georgiou v. Comm'r of Dep't of Indus. Accidents*, 67 Mass. App. Ct. 428, 435 (2006) (records of employees who had been involved in an accident which placed them on disability leave).

In contrast, the Records here are like a large puddle that is three inches deep. The Records reveal so little as to each person that their disclosure poses little, if any, privacy concern. When serious

privacy issues arise, they tend to arise primarily from "deep" datasets that have many attributes assigned to a particular individual, see Micah Altman et al., *Towards a Modern Approach to Privacy-Aware Government Data Releases*, 30 Berkley Tech. L.J. 1967, 2037-2040 (2016), or with datasets that contain information that is itself inherently sensitive, such as a social security number.

Within the birth index, each entry reveals only a person's first and last name, date of birth, and parents' names (along with the location of the record in the locked vault). See *Globe Br. 7*. Within the marital records, each record reveals only the person's first and last name, date of marriage, name of spouse, and place where the license was filed (along with a certificate number and the location of the record in the locked vault). *Id.* at 7-8. This is more akin to the "shallow" records this Court has considered previously, and ordered released. See *Cape Cod Times*, 443 Mass. at 587-588 (names and addresses of "reserve deputy sheriffs"); *Pottle*, 395 Mass. at 863 (list of employee absentee records).

**B. These records do not pose a privacy concern, despite being "broad."**

The only substantial difference between this case and cases in which this Court required release of "shallow" datasets is that the Records are "broader," that is, they contain information about a larger number of persons. There are distinct issues courts may wish to consider regarding disclosing broad datasets, but in this case, there is no increased privacy concern.

On one level, broader datasets can provide greater insights because of one's ability to perform more advanced statistical analysis, thus increasing their public utility, and by extension the public benefit in their disclosure. Analyzing broad datasets can help discover correlations between the data that a human observer might otherwise miss. See Daniel T. Larose & Chantal D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining* 8-14 (2d ed. 2014) (describing the ability to use larger datasets to help classify records within a set, find associations in data, and predict future outcomes). To use a hypothetical example, an analysis of the Records may show that people born in Fitchburg are two times

more likely to have a name that starts with the letter A than those born in Gardner, or those who file for a marriage license in Cambridge are 30% more likely to hold a wedding on a Saturday in June. Such correlations can be quite useful, but they would not disclose additional or more invasive information about any individual. See *id.* at 6-7.

Where there is an increased privacy risk in broad datasets, it tends to be with datasets that are also "deep" and anonymized in some form. This is primarily because some scholars have now developed the ability to "de-anonymize" some complex datasets by combining them with other information that can reveal individual identities. This is typically true when the anonymized data uses combinations of quasi-identifiable information, such as zip codes combined with age and gender, which wind up being unique enough to return a researcher back to individual people. See, e.g., Boris Lubarsky, *Re-Identification of "Anonymized" Data*, 1 *Geo. Tech. L. Rev.* 202 (2017). When faced with an anonymized dataset, courts may decide to consider whether data is safe for disclosure only because it is not personally identifiable, and carefully examine whether one could re-identify the data set. See

generally Giske Ursin et al., *Protecting Privacy in Large Datasets - First We Assess the Risk; Then We Fuzzy the Data*, 26 *Cancer Epidemiology, Biomarkers and Prevention* 1219, 1220 (2017).

But this case is far simpler. The records here are not deep and anonymized; they are shallow and identified. This Court need not run through the hypothetical question of whether an anonymous but sensitive set of data can be identified. *Cf. Globe Newspaper Co. v. Boston Ret. Bd.*, 388 Mass. 427. It need only examine the Records as they are, and in this case, the Records are not sensitive.

**IV. RELEASE OF STATEWIDE VITAL STATISTICS RECORDS HAS NUMEROUS PUBLIC BENEFITS.**

Even if the Superior Court correctly concluded that the Records objectively contain intimate details of a highly personal nature, it erred in holding that the public interest in obtaining the information did not substantially outweigh the seriousness of the invasion of privacy.

The Superior Court overlooked the numerous potential beneficial uses of the Records. Datasets like the Records can be used to improve government accountability, gain new insights about systemic

health and welfare of Commonwealth residents, and inform local news reporting.<sup>8</sup>

Journalism based on the Records can help improve government accountability. As governments have taken existing analog records and converted them to a digital format, news organizations have developed teams that can quantitatively analyze the information and report on its insights, giving rise to a form of journalism known as “data journalism.” See generally D. Victoria Baranetsky, *Data Journalism and the Law* (Sept. 19, 2018), Tow Ctr. for Digital Journalism, [https://www.cjr.org/tow\\_center\\_reports/data-journalism-and-the-law.php/](https://www.cjr.org/tow_center_reports/data-journalism-and-the-law.php/) (“Since about 2008, the explosion of data journalism – defined as journalism that heightens the role numerical information plays in storytelling – is now a driving force in newsrooms around the country.”); *All Data Sets*, ProPublica,

---

<sup>8</sup> The Records can also help improve the overall quality of journalism as well. The Superior Court did not agree that the ability to verify identities with public records provides a public benefit, but this is a common practice in responsible reporting that would be more prevalent with access to this dataset. See generally Mark Memmott, *‘Memmos’: Memmott’s Missives & Musings*, NPR Ethics Handbook, (Nov. 20, 2014), <http://ethics.npr.org/category/memos-from-memmott/#1984> (NPR’s Standards & Practices Editor, in giving advice to radio reporters, “it’s wise to check public records”).

<https://www.propublica.org/datastore/datasets>  
(collecting various government datasets for analysis).  
A 2017 study revealed that 42 percent of surveyed  
journalists "said they use data regularly to tell  
stories, and 51 percent of news organizations have a  
dedicated data journalist on staff." Anne Glover &  
David Beard, *Study Shows Data Reporting Gaining Hold  
in Newsrooms*, Poynter (Sept. 18, 2017),  
[https://www.poynter.org/tech-tools/2017/study-shows-  
data-reporting-gaining-hold-in-newsrooms/](https://www.poynter.org/tech-tools/2017/study-shows-data-reporting-gaining-hold-in-newsrooms/).

Using vital statistics data, journalists can hold  
governments to account for their actions. For example,  
Los Angeles journalists used death records of  
California citizens to show that votes were being cast  
from voters long after their death, raising election  
integrity concerns. David Goldstein, *CBS2  
Investigation Uncovers Votes Being Cast From Grave  
Year After Year*, CBS Los Angeles (May 23, 2016),  
[https://losangeles.cbslocal.com/2016/05/23/cbs2-  
investigation-uncovers-votes-being-cast-from-grave-  
year-after-year/](https://losangeles.cbslocal.com/2016/05/23/cbs2-investigation-uncovers-votes-being-cast-from-grave-year-after-year/). The public records and journalism  
organization MuckRock has been using sets of public  
records to expose nationwide issues around maternal  
deaths during childbirth, especially in rural areas,

raising state healthcare administration and oversight concerns. Adanya Lustig, *In Rural Hospitals, Women are Dying During Childbirth – and States Aren't Keeping Track*, MuckRock (Dec. 4, 2017), <https://www.muckrock.com/news/archives/2017/dec/04/maternal-deaths-map/>. The public television series Frontline exposed ongoing instances of child marriage across the country using sets of public records, which led the State of Delaware this year to become the first to ban the practice. Anjali Tsui, *Delaware Becomes First State to Ban Child Marriage*, Frontline (May 9, 2018), <https://www.pbs.org/wgbh/frontline/article/delaware-becomes-first-state-to-ban-child-marriage/>.

Beyond questions of policy and law enforcement, datasets like the Records can be used by both journalists and researchers to develop a better understanding of societal issues and population trends in Massachusetts. The Society of Professional Journalists notes that vital statistics can be used to “track population and demographic trends,” while marriage records, specifically, can be used to “monitor marriage statistics or confirm marriage

data." *FOI A to Z*, Society of Professional Journalists, <https://www.spj.org/opendoors7.asp>.

Both death and birth records are viewed as crucial resources for public health research and journalism, and "[t]he many attempts to make birth and death records a secret shared only with a select few do real damage to efforts to improve health in communities large and small." William Heisel, *Well Sourced: Use Vital Records to Find Patients, Document Trends*, Ctr. for Health Journalism (May 5, 2015), <https://www.centerforhealthjournalism.org/2015/05/04/well-sourced-use-vital-records-find-patients-document-trends>. Using these records, journalists and researchers can look broadly at state-wide trends such as average lifespans, while also having the ability to narrow the scope of reporting to county-by-county trends. The same can be said for marriage records, as this collection of records can tell reporters and researchers more about localized trends in marriages. See, e.g., Steph Yin, *When Did Americans Stop Marrying Their Cousins? Ask the World's Largest Family Tree*, N.Y. Times (Mar. 1, 2018), <https://www.nytimes.com/2018/03/01/science/cousins-marriage-family-tree.html> (tracing the brief uptick and then decline of cousins

marrying over the mid-to-late 19th Century, using a large private dataset of vital records).

Finally, journalists from various news organizations in the region, including *amici* here who are student journalists, can use the data from these datasets to advance their own coverage of university-based issues. For example, the numerous student newspapers located in and around Boston, Amherst, Worcester, and elsewhere in Massachusetts can use the marriage records in Massachusetts to compare trends in "spousal hires" of faculty among at different universities, or look to see what proportions of faculty are Massachusetts natives, or develop quantitative comparisons of legacy admissions. This type of reporting is essential in a state that has such a high concentration of institutions of higher education. Without these datasets requested here, these stories may not be possible.

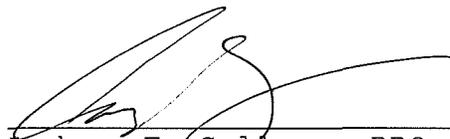
And it is precisely with student journalists, as well as smaller news organization and freelancers, where the necessity of obtaining the records in bulk is most apparent, as these reporters may not have the time to travel to an office location, and simply cannot afford access to these records on a pay-per-

hour basis, as the Department presently provides. See generally *Globe Br.* 4-5; see also Michael Barthel, 5 *Facts About the State of the News Media in 2017*, Pew (Aug. 21, 2018), <http://www.pewresearch.org/fact-tank/2018/08/21/5-facts-about-the-state-of-the-news-media-in-2017/> (noting that while revenues for cable news have grown, newspaper and local TV news revenue declined in 2017); *Untold Stories: A Survey of Freelance Investigative Reporters*, ProjectWorld, 15-18 (2015), [http://www.firenewsroom.org/sites/default/files/PW\\_Freelancer\\_Survey\\_2015.pdf](http://www.firenewsroom.org/sites/default/files/PW_Freelancer_Survey_2015.pdf) (explaining that freelancer reporters are often hindered by lack of access because of expensive tools). In an age of declining newsroom revenues, student newsrooms, smaller and nonprofit outlets, and freelancers often do not have the same resources that other news organizations have to obtain public records, and their important insights into these records are lost if they are only made available to those with time and money.

### **Conclusion**

For the reasons stated above, *amici* respectfully urge this Court to reverse the judgment of the Superior Court, and find that the records in question may not be withheld under Exemption (c).

Respectfully submitted,



Andrew F. Sellars, BBO #682690  
**BU/MIT Technology & Cyberlaw Clinic**  
Boston University School of Law  
765 Commonwealth Ave.  
Boston, MA 02215  
Tel: (617) 358-7377  
Fax: (617) 353-6944  
sellars@bu.edu

*Counsel for amici curiae*<sup>9</sup>

---

<sup>9</sup> *Amici* wish to thank Boston University School of Law students Alexandra Faustin, Zachary Sisko, Lyndsey Wajert, and Patrick Wilson for their invaluable contributions to this brief.

**Certification Pursuant to Mass. R. App. P. 16(k)**

I hereby certify that this brief complies with the rules of this Court pertaining to the filing of briefs, including, but not limited to: Mass. R. App. P. 16(a)(6) (pertinent findings or memorandum of decision); Mass. R. App. P. 16(e) (references to the record); Mass. R. App. P. 16(f) (reproduction of statutes, rules, regulations); Mass. R. App. P. 16(h) (length of briefs); Mass. R. App. P. 18 (appendix to the briefs); and Mass. R. App. P. 20 (form of briefs, appendices, and other papers).

  
\_\_\_\_\_  
Andrew F. Sellars

**Certificate of Service**

I, Andrew Sellars, hereby certify that on December 21, 2018, I served two copies of the above brief on the following counsel:

Jonathan Albano  
Morgan, Lewis & Bockius LLP  
One Federal Street  
Boston, MA 02110-1726

William W. Porter  
Assistant Attorney General  
Office of the Attorney General  
One Ashburton Place  
Boston, MA 02108

  
\_\_\_\_\_  
Andrew F. Sellars