November 1, 2018

Honorable Tani Cantil-Sakauye, Chief Justice
Honorable Associate Justices
Supreme Court of California
350 McAllister Street
San Francisco, California 94102

Re: Letter of Amicus Curiae Electronic Frontier Foundation in Support of Petition for Review: *Sander v. State Bar of California* (2018) 26 Cal.App.5th 651; Supreme Court of the State of California Case No. S251671

Dear Chief Justice Cantil-Sakauye and Associate Justices of the Court:

The Electronic Frontier Foundation ("EFF") submits this letter in support of the Petition for Review filed by Richard Sander and the First Amendment Coalition in the above-captioned case, *Sander v. State Bar of California*. In accordance with California Rule of Court 8.500(g)(1), a copy of this letter was served on all parties to the case.

EFF urges this Court to grant the Petition for Review because the Court of Appeal decision rewrites the California Public Records Act ("CPRA") in a way that could limit the ability for Californians to access the vast amount of public data that state and local agencies are generating on our behalf. The Petition thus raises "an important question of law" that this Court must settle: does anonymization of public data amount to a creation of new records under the CPRA? (R. of Ct. 8.500(b)(1).) Because the CPRA and California Constitution require that public access be construed broadly, and limits on access be construed narrowly, the answer should be no.[1] Moreover, the Court of Appeal's holding that anonymizing public data creates a new record grafts a judicial exception on to the CPRA, which frustrates both the California Legislature's intent in enacting the CPRA and the public's right of access to records.

[1] In relevant part, Article I, Section 3 of the California Constitution provides:

> A statute, court rule, or other authority, including those in effect on the effective date of this subdivision, shall be broadly construed if it furthers the people's right of access, and narrowly construed if it limits the right of access[.]

Cal. Const., art. I, § 3, subd. (b), par. (2).

This Court's intervention is necessary to prevent the decision from blunting the CPRA's application to the largest and most rapidly growing set of public records in California: electronic data. As explained below, state and local agencies are increasingly collecting or receiving vast amounts of data, a reality in light of our digital world. Given the enormous increase in the volume of records that government can maintain, public oversight of government's activities through tools like the CPRA becomes all the more necessary. Should the Court of Appeal's decision stand, it could render the CPRA less relevant to public access and oversight, and it could foreclose access to public data that contains identifying or personal information, no matter how strong the public interest in such data may be. This would thwart Californians' efforts to learn about important, and sometimes controversial, government activity. EFF thus respectfully urges this Court to grant the Petition for Review.

Moreover, the Court's review is necessary "to secure uniformity of decision." (R. of Ct. 8.500(b)(1).) In contrast to the Court of Appeal decision here, the Contra Costa County Superior Court recently held de-identification of data sought under the CPRA does *not* create a new record, much less permit the agency to withhold it under a "new records" exemption not found in the text of the CPRA. (*Exide Tech. v. Cal. Dept. of Public Health* (Contra Costa County. Super. Ct. April 13, 2018, No. N16-0737), petn. for writ of mandate denied (Case No. A154209 (1st App. Dist. May 30, 2018)) (hereafter *Exide*).) While this opinion is not precedential, it explicitly relies on this Court's guidance from *ACLU Foundation of Southern California v. Superior Court* (2017) 3 Cal.5th 1032 (hereafter *ACLU*) that the trial court should "consider on remand 'the feasibility of, and interests implicated by, methods of anonymization petitioners have suggested[,]' and to explore 'other methods of anonymization and redaction as well.'" (*Exide*, slip op. at 21 (quoting *ACLU*, 3 Cal.5th at 1046–1047).) *Exide* demonstrates the error of the Court of Appeal's decision here and shows that the issue of de-identifying data in response to CPRA requests is prominent in California courts.

Interest of Amicus Curiae

EFF is a San Francisco-based non-profit civil liberties organization that has worked for more than 25 years to protect and promote fundamental liberties in the digital world. As part of its mission, EFF has challenged the over-withholding of government records by California state agencies, most recently in its case, brought with the ACLU Foundation of Southern California, challenging the Los Angeles Police and Sheriff's Departments' withholding of license plate records under the CPRA. (*See ACLU*, 3 Cal.5th.) EFF has also served as amicus in other cases that seek to uphold the public's right to access government records under the CPRA and the federal Freedom of Information Act (FOIA), including *City of San Jose v. Superior Court* (2017) 2 Cal.5th 608 and *Sierra Club v. Superior Court* (2013) 57

Cal.4th 157. Amicus will assist the Court by demonstrating that this case has consequences for the public's ability to access electronic records and government data.

INTRODUCTION

In this case, the Court of Appeal held that applying anonymization protocols to State Bar data constitutes the creation of a new record. This holding is both out of step with the realities of modern data collection and storage and fails to comport with the CPRA's constitutional mandate to broadly construe statutory disclosure requirements to further the people's right of access. This Court should grant review to ensure the CPRA stays relevant in the age of databases and digital records.

Modern governments generate and consume vast amounts of digital data about members of the public. This implicates two fundamental rights, both explicitly recognized by the California Constitution. First, the public has a right to access this data. This right enables the public to understand what its government is doing and to use the data to expose government inefficiency or malfeasance. Second, the people described by the data have a right to privacy. When digital data are made public without protecting people's privacy, others could expose intimate details of their lives to unwanted scrutiny. These two fundamental rights must be carefully balanced.

Across the country, data custodians are innovating ways to share digital data in a manner that advances access to information without invading privacy. A critical method is known as anonymization (or de-identification). This means that before data custodians release a dataset, they apply techniques to remove or obscure information that could be used to identify the people described by the data. A growing field of scholarship is creating best practices to de-identify data, and government agencies in California are already using anonymization techniques.

In the case at bar, however, the appellate court held the State Bar does not have to apply anonymization techniques to the records sought by petitioners because to do so would require the state to create a "new record," which, the court held, was not mandated by the CPRA. This holding is incorrect and violates both the spirit of the CPRA and the letter of the law. Anonymization is simply a modern way of redacting exempt information from otherwise non-exempt records that the CPRA requires government to release. And it is a process that can protect bar members' privacy interests while still allowing for public disclosure of important state-held data. Anonymization and other sophisticated privacy-protecting techniques are necessary to ensure the proper balance between the competing interests of government transparency and individual privacy, as the CPRA and the California Constitution require. (*City of San Jose,* 2 Cal.5th at 616.)

This Court recently ruled: "Our case law recognizes that the CPRA should be interpreted in light of modern technological realities." (*ACLU*, 3 Cal.5th at 1041.) But the appellate court's cramped interpretation of the CPRA goes against this mandate. It is likely to have far-reaching consequences that frustrate access to vast amounts of government digital data in which the public has a legitimate interest. This Court should grant review to correct these errors.

ARGUMENT

I. With the Explosive Growth of Government Data, the CPRA Must Be Interpreted to Provide Broad Access to Data-Rich Records

With the growth of Internet-enabled technologies, the amount of data created worldwide each year continues to surpass the years before.[2] Government data has followed this trend as state agencies and local governments have invested in data collection, utilization, and management and are finding ways to digitize older records to make them more useful to the public. As data and data-driven algorithms increasingly become an integral part of how government agencies function, public access to data is essential to government accountability and oversight. Although the CPRA was enacted before electronic records were as prevalent as they are today, the legislature has continued to update the law to ensure it allows access to electronic records and data sets. Local agencies in California have also established policies that promote access to some repositories of government data. Further, some agencies have already developed anonymizing practices that promote transparency while protecting privacy.

These policies can and should inform how agencies process individual requests under the CPRA. In this age of explosive data creation and utilization, it would be antithetical to the original purpose of the CPRA to allow government agencies to shield their datasets and data practices from the public, merely by refusing to grapple with the consequences of creating sensitive data about identifiable people that are also public records. Instead, agencies must meaningfully evaluate and adopt new technologies that will allow them to release records in ways that properly balance privacy and transparency.

[2] *See* Åse Dragland, *Big Data, for better or worse: 90% of world's data generated over last two years*, Science Daily (May 22, 2013) https://www.sciencedaily.com/releases/2013/05/130522085217.htm. All websites last visited on January 29, 2018.

A. As Modern Data Capabilities Have Grown, Agencies Are Finding New Ways to Open Up Access to Their Records, Including Records Containing Highly Sensitive and Private Data

1. *State and Local Governments Generate Vast Amounts of Data Each Year*

State and local agencies are producing and collecting data at an unprecedented rate. As part of an effort to make the increased data collection more transparent, the California Legislature passed Senate Bill 272 in 2016, adding section 6270.5 to the Government Code.[3] The law requires local agencies, with the exception of school districts, to post catalogs of their databases in a "prominent location" on their websites and to make this data readily available to the public. (Gov. Code § 6270.5(a).)[4] These catalogs are meant to disclose all the data systems an agency uses as a primary source of records or to collect information about the public. The legislature found that by "turning internally gathered and maintained data into usable information for the public to access," California agencies allow the public to "leverage [this data] for the benefit of their communities." (Sen. Bill No 272 (2015–2016 Reg. Sess.) § 1, subd. (d).) These catalogs have not only become important government records in their own right, they can serve as a menu of records that members of the public may request under the CPRA.

As cities, counties, and individual agencies have put their database catalogs online, the public has learned just how many data systems they each maintain.[5] Each agency's catalog may include a handful to several hundred databases, which in turn include countless records. For example, the City and County of San Francisco lists 464 individual data systems, with approximately half of the systems

---

[3] *See* Sen. Bill No. 272 (2015–2016 Reg. Sess.) *available at* https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201520160SB272.

[4] Except where noted, all statutory citations are to the Government Code.

[5] EFF lists available government digital datasets in a single linkable document that now includes 442 California local agencies. (*See California Database Catalogs 2016*, Electronic Frontier Foundation, https://www.eff.org/pages/california-database-catalogs-2016.)

updated with new data either daily or continuously.[6]

It is not surprising that cities and counties maintain so many datasets, because so much of the information agencies rely on to do their work has either been digitized or is now digital throughout its lifecycle. For example, many counties and cities have digitized older vital and official records, such as birth and death certificates and property records, to make them searchable, both for the benefit of agency employees and for the general public.[7] The state has also taken steps to offer more services to the public online, such as the Department of Motor Vehicles' website that allows Californians to renew their vehicle registration or driver's license and update their address information online.[8] Similarly, the Secretary of State has an online portal that allows businesses to file their required financial statements electronically.[9] These services not only allow the public to provide information to the state and conduct state business from the comfort of their own home, they also allow the public to search through the information provided to the state by others. Digital services, and transparency around those services, create increased efficiencies for connecting the public with the day-to-day operations of their government.

2. *State Agencies Are Finding Ways to Make Sensitive Data Available to the Public While Still Protecting Privacy*

As more and more data is collected and stored digitally, California agencies are finding ways to make even highly sensitive information available to the public

---

[6] *Inventory of citywide enterprise systems of record*, DataSF (last updated Jul. 2, 2018) https://data.sfgov.org/City-Management-and-Ethics/Inventory-of-citywide-enterprise-systems-of-record/ebux-gcnq/data.

[7] *See, e.g.*, *2017 was a busy year for the Recorder-Clerk*, The ARC Blog (Jan. 10, 2018) https://sbcountyarcblog.org/2018/01/11/2017-busy-year-recorder-clerk (noting the San Bernardino Recorder's Office digitized vital records, such as birth and death certificates, dating back to 1910, and property records, dating back to 1958); Emily Alpert Reyes, *Many L.A. building records now just a few clicks away*, L.A. Times (Jun. 18, 2015) http://www.latimes.com/local/lanow/la-me-ln-online-building-records-20150618-story.html (noting the Lost Angeles Department of Building and Safety digitized more than 13 million records dating back to 1905).

[8] *See* California Department of Motor Vehicles, Online Services, https://www.dmv.ca.gov/portal/dmv/detail/online/onlinesvcs.

[9] *See* California Secretary of State, Bizfile California, http://www.sos.ca.gov/business-programs/bizfile.

without disclosing the identity of the people described by that information. For example, the California Health and Human Services Agency (CHHS) has established de-identification guidelines to allow public access to certain datasets while still complying with all laws and patient protections, including the de-identification requirements in the federal Health Insurance Portability and Accountability Act.[10] The CHHS open data portal provides de-identified records on the rates of certain diseases, such as the number of cases of antibiotic-resistant staph infections for each hospital in a given year;[11] the demographics of individuals who access healthcare services, including applicants for insurance affordability programs by country of origin;[12] and how transportation and other environmental factors may impact a person's health, including a comprehensive database of motor vehicle accidents leading to death or serious injury, with the individual incidents and races of the victims given a numerical value to prevent victim re-identification.[13]

The de-identification guidelines published by CHHS advocate for a case-by-case approach when creating an open dataset. In doing so, the agency recognizes that individual record-level data has a greater probability of including sensitive information than anonymized summary-level data.[14] This case-by-case approach increases the possibility for appropriately balancing access and privacy, as exemplified by the motor vehicle accident database described above that substitutes sensitive personal information with an anonymous value.

Similarly, the City of San Francisco has created a privacy-protective balancing test to determine how city data should be released to the public.

---

[10] *Data De-identification Guidelines (DDG)*, California Department of Health Care Services, at 5 (Nov. 22, 2016) http://www.dhcs.ca.gov/dataandstats/Documents/DHCS-DDG-V2.0-120116.pdf.

[11] *Methicillin-resistant Staphylococcus aureus bloodstream Infections (MRSA BSI) in Healthcare*, CHHS Open Data (2013–2017) https://data.chhs.ca.gov/dataset/methicillin-resistant-staphylococcus-aureus-bloodstream-infections-mrsa-bsi-in-healthcare.

[12] *Ethnicity of Applicants for Insurance Affordability Programs*, CHHS Open Data (2016–2018) https://data.chhs.ca.gov/dataset/dhcs_ethnicity-of-applicants-for-insurance-affordability-programs.

[13] *Road Traffic Injuries*, CHHS Open Data (2002–2010) https://data.chhs.ca.gov/dataset/road-traffic-injuries-2002-2010.

[14] *See Data De-identification Guidelines*, *supra*, at pp. 20, 43.

Following a 2009 directive,[15] all agencies must publish the data they collect and process on a single website for the public to access. Also, the city created a rubric to prioritize the publication of highly requested data, even where some privacy concerns may exist, before less publicly interesting data.[16] Further the city's Open Data Release Toolkit, which provides steps on how to release data, suggests masking identifying variables, like name or ID number; obscuring quasi-identifying information, like race and age; and undertaking other methods of de-identification when appropriate.[17]

The guidelines and tests discussed above, along with the agencies' initiatives to make datasets publicly available in an open-source format, show that government agencies are able, at every stage of the data management process, to balance the interests of individual privacy with the community's interest for access.

However, while datasets published through these programs have been prioritized for public accessibility, many other datasets, like the state bar's database at issue here, contain records that are unquestionably of interest to the public but are not maintained in a way that would allow for automatic release without risking severe privacy harm. The public should not have to wait for government agencies to voluntarily modify and publish those datasets online. Instead, state agencies must also find ways to disclose the data in response to individual public records requests.

B. Releasing Public Data Increases Government Oversight

The California Supreme Court squarely recognized the strong public interest in access to state bar data, holding "it seems beyond dispute that the public has a legitimate interest in whether different groups of applicants, based on race, sex or ethnicity, perform differently on the bar examination and whether any

---

[15] S.F. Mayor Gavin Newsom, *San Francisco Government and Technology: How We're Innovating*, Mashable (Oct. 21, 2009) http://mashable.com/2009/10/21/san-francisco-government/#kX_VXug3cOqp (reporting San Francisco's "effort to improve access to city data led to the creation of new services that are now featured in the DataSF App Showcase"). San Francisco's Open Data Portal can be found at https://datasf.org/.

[16] Former S.F. Chief Data Officer Joy Bonaguro, *How to Unstick Your Open Data Publishing*, DataSF (last updated Jul. 7, 2015) https://datasf.org/blog/how-to-unstick-data-publishing.

[17] *Open Data Release Toolkit: Privacy Edition v1.2*, DataSF (Nov. 3, 2016) https://docs.google.com/document/d/1MhvEuGKFuGY2vLcNqiXBsPjCzxYebe4dJicRWe6gf_s/edit#heading=h.v77s0yo7ojk7.

disparities in performance are the result of the admissions process or of other factors." (*Sander v. State Bar of Cal.,* (2013) 58 Cal.4th 300, 324.) Access to the granular data contained in the state bar's databases—even in anonymized form—would allow the public to evaluate government activity on a much deeper level than mere access to summary data about the bar's programs.

The *Exide* case supports this analysis. In a challenge to the California Department of Public Health's refusal to release de-identified records of lead poisoning in Los Angeles County, the superior court found that records showing the source of lead exposure were "vital not only for defining the problem but also for identifying the measures needed to address it." (*Exide*, slip op. at 16.) The court recognized the state legislature's finding that "knowledge about where and to what extent harmful childhood lead exposures are occurring in the state could lead to the prevention of these exposures, and to the betterment of the health of California's future citizens." (*Id.* at 10.) The court further found that the existing publicly-provided information about blood lead levels was "insufficient to meaningfully analyze these important issues." (*Id*. at 16.) For these and other reasons, the public interest in disclosure of the data outweighed the privacy interests in refusing to release it in any form.

In another example, before CHHS committed to provide hospital data through its open data portal, investigative researchers at ProPublica filed CPRA requests seeking records of all HIPPA violations and citations issued by the state.[18] ProPublica found inconsistent enforcement by the state and a large disparity between HIPPA violations that hospitals and other entities were reporting and citations that were then issued.[19] Moreover, Los Angeles County, which is home to over 100 hospitals, received only a handful of violations despite very public data-breaches at a number of hospitals.[20]

Through the CPRA, ProPublica requested specific fields designed to protect the privacy of patients: hospital name, hospital ID, start date of survey, end date of

---

[18] Charles Ornstein, *The Consequences for Violating Patient Privacy in California? Depends Where the Hospital Is*, ProPublica (Dec. 31, 2015), https://www.propublica.org/article/california-patient-privacy-law-inconsistent-enforcement.

[19] *Id*.

[20] *Id*.

survey, survey ID, and details of any deficiencies.[21] They turned the records received in response to their CPRA request and a FOIA request to the federal Health and Human Services into a consumer-friendly database called "HIPPA Helper" where people can look up specific hospitals, pharmacies, or types of violations to find HIPPA violation reports and make informed decisions about their healthcare providers.[22] The California Department of Public Health, a component of CHHS, scrubbed records of identifying patient and employee information, but provided valuable details about incidents[23] — like a violation at Eisenhower Medical Center where an employee in the billing department looked up the age and marital status of seven different patients in an attempt to find a potential date for a friend.[24]

Access to the state's education data, specifically, is vital for Californians to make informed decisions about where to apply and attend institutions for higher education. Relying on de-identified data received from public records requests to three California law schools, Mr. Sander has theorized that when some minority law school applicants attend schools where the median test scores are above what they scored, they do not perform as well on the state bar exam as other minority law school applicants who attended schools where median test scores more closely matched their own.[25] Access to additional similar data would allow for important independent analysis of Mr. Sander's research.

---

[21] Charles Ornstein and Annie Waldman, *Methodology: How We Analyzed Privacy Violation Data*, ProPublica (Dec. 29, 2015), https://www.propublica.org/article/methodology-how-we-analyzed-privacy-violation-data.

[22] Charles Ornstein, et. al., *HIPPA Helper*, ProPublica (Dec. 29, 2015), https://projects.propublica.org/hipaa/.

[23] *See e.g.,* Report, Scripps Mercy Hospital, Report ID: O0N211.03, *available at* https://projects.propublica.org/hipaa/reports/876f85dba4fba9dc28d7362381d26f5b37478e51; Report, San Francisco General Hospital, Report ID: 0J8011.02, *available at* https://projects.propublica.org/hipaa/reports/3dc9ad378e731c20affca233f89c7cd623ae160d; Report, University Of California Medical Center, Report ID: WZ5T11.01, *available at* https://projects.propublica.org/hipaa/reports/7e42b462480627069d26c092a8ccb311be6b2a17.

[24] Report, Eisenhower Medical Center, Report ID: BVAW11, *available at* https://projects.propublica.org/hipaa/reports/d402f53a3dc375aa4ac7d51d0f3a02b3ca22e584.

[25] Richard Sander and Robert Steinbuch, *Mismatch and Bar Passage: A School-Specific Analysis*, UCLA School of Law, Public Law Research Paper No. 17-40 (Oct. 17, 2017) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3054208.

Similarly, data that the state has made public about student performance at community colleges has helped to change how community colleges assess who can enroll in transfer-level math and English courses.[26] Data showed that black and Latino students were being placed in remedial, nontransferable courses in much higher numbers than white students, often based solely on results of an entrance exam. In part because these courses extended the amount of time a student had to attend community college before they could transfer to a four-year college, the dropout rates were much higher for students placed in these courses than for those placed directly into transferrable courses. The inequalities resulting from the community college remedial placement process led the California legislature to pass a law in 2017 requiring community colleges to "use, in the placement of students into English and mathematics courses . . . one or more of the following: high school coursework, high school grades, and high school grade point average." (Assem. Bill No. 705 (2017–2018 Reg. Sess.).)[27] It also prohibited community colleges from requiring students to enroll in remedial English or math that would lengthen their time to complete a degree unless placement research showed "that those students are highly unlikely to succeed in transfer-level coursework." (*Id.*) The goal of these changes is to address the inequity of current practices by "maximiz[ing] the probability that the student will enter and complete transfer-level coursework in English and mathematics within a one-year timeframe." (*Id.*)

More broadly, public access to government data holds the potential to provide more accountability and to identify ways state and local agencies can better serve the public. In light of the government oversight and public benefit that access to data allows, agencies have an obligation to the public to make the data available in some form. Agencies can do this in ways that minimize harms to individual privacy by following the guidelines set out in the state's open data programs or by adopting anonymization protocols.

---

[26] Meredith Kolodner, et. al., *Remedial classes: A community college 'segragation machine,* inewsource (Dec. 13, 2017), https://inewsource.org/2017/12/13/california-remedial-community-college.

[27] Assem. Bill No. 705 (2017-2018 Reg. Sess.) *available at* https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB705.

II. This Court Should Grant Review to Correct the Court of Appeal's Erroneous Decision that Anonymization of Data Results in the Creation of New Records.

The Court of Appeal's holding that anonymizing the data sought in this case would violate the CPRA because it requires the creation of new records is wrong for two reasons. First, neither the plain language of the CPRA nor the case law cited by the Court of Appeal supports its position. Second, the protocols proposed by Petitioners do not require the State Bar to create new records. Petitioners' efforts to anonymize the data at issue in this case are equivalent to redacting otherwise exempt information from records, and thus well within the bounds of the CPRA.

A. The CPRA Does Not Contain a "New Records" Exemption

The *Exide* case demonstrates the plain error of the Court of Appeal's holding that the CPRA does not require agencies to create new records. In *Exide,* the court rejected the argument that de-identifying data subject to a CPRA request would create a new record in violation of the CPRA. (*Exide*, slip op. at 19–21.) Beginning with the text of the CPRA, the court recognized that the CPRA only permits withholdings that are expressly authorized by the statute. (*Id*. at 19–20.) The court also recognized the text of the CPRA contains no express or implied "new records" exemption. (*Id.*) It does not include any limitations on agencies' duties to extract information from public records, to manipulate information within records, or to withhold exempt information while making non-exempt information available to the public. (*Id.* at 20.) Thus, in light of the plain language of the CPRA and the fact that there "is no express provision of the CPRA exempting 'new records,'" the *Exide* court held that the agency could not withhold the records on that ground. (*Id*. at 19.)

The *Exide* court went on to hold that, in light of the CPRA and California Constitution's mandates that the CPRA be construed liberally to promote access, while construing limitations on access narrowly, it could not "create new exemptions to disclosure that are not embodied in the express provisions of the CPRA." (*Id*. at 20.)

Further, the *Exide* court recognized that creating a "new records" exemption would conflict with other CPRA provisions that expressly require agencies to provide electronic records to requesters, including producing copies even when the "request would require data compilation, extraction, or programming." (*Id*. (quoting § 6253.9(b).) The CPRA goes further, and expressly anticipates that such extraction and manipulation will be necessary in some cases and that agencies may pass along to requesters the corresponding costs "to construct a record, and the cost of programming and computer services necessary to produce a copy of the record."

(§ 6253.9(b).) Likewise, such extraction and manipulation are contemplated by the CPRA's rule that agencies may delay their responses based on the "need to compile data, to write programming language or a computer program, or to construct a computer report to extract data." (§ 6253(c)(4).) In light of the statute requiring agencies to take such steps to make public data available, the court concluded that "[t]hese provisions of the CPRA cannot be reconciled with Respondent's assertion that public agencies may withhold information in public records if they require creation of a 'new record.'" (*Exide*, slip op. at 20.) The *Exide* court thus recognized that creating a "new record" exemption to the CPRA would render section 6253.9 irrelevant, violating a basic canon of statutory construction. (*Id.*; *see also South Carolina v. Catawba Indian Tribe, Inc.* (1986) 476 U.S. 498, 510 n. 22 (collecting cases that hold courts must interpret statutes to give full effect to them and not adopt interpretations that render portions meaningless or superfluous).)

The Court of Appeal's decision also erred in relying on cases such as *Fredericks v. Superior Court* (2015) 233 Cal.App.4th 209 to justify its position that agencies cannot be required under the CPRA to create new records from their existing data and information. *Fredericks* does not support this proposition and is inapplicable to this case because its discussion about agencies' duties under the CPRA was dicta. *Fredericks* addressed the scope of an agency's duties to extract information from law enforcement investigative files pursuant to § 6254(f)(2), a statute that specifically requires agencies to extract and produce certain delineated categories of information from police complaints or requests for assistance. (*Id.* at 215.) *Fredericks'* statement—that requiring agencies to create new records in response to a public records request would exceed the agencies' statutory duties under the CPRA—was dicta because the court never had to address whether the petitioner's request would require the agency to create a new record. (*Id.* at 227.)[28]

B. Anonymizing Data Does Not Result in the Creation of a "New Record"

Even if this Court were to agree with the Court of Appeal that the CPRA does not require agencies to create new records, that point is moot because the privacy protocols proposed by Petitioners do not require the State Bar to create new records. Instead, like the protocols upheld by the court in *Exide*, they require the Bar to manipulate existing public records to produce data in a format that serves the public interest in government transparency (by disclosing non-exempt information) while at the same time protecting the privacy interests of state bar applicants (by withholding exempt information). Again, the CPRA explicitly

[28] *Sander* also does not involve extracting information from investigative files, so *Fredericks* is not on point.

contemplates such data manipulation by allowing agencies to charge requesters reasonable costs for producing records that require "data compilation, extraction, or programming." (§ 6253.9(b)(2).)

Other cases, involving both the CPRA and the federal Freedom of Information Act, also support the need for this Court to review the Court of Appeal's decision because they are more closely aligned with Petitioner's request here than the cases it relied upon. For example, this Court recently recognized in *ACLU* that agencies may be required to manipulate their data so that they can release records while still protecting privacy. The Court stated, "While real parties may not have designed their system to facilitate CPRA disclosure as a 'native function,' randomizing [data] or deleting columns from a spreadsheet, for example, would seem to impose little burden." (*ACLU*, 3 Cal.5th at 1047.) The *Exide* court found that this Court's remand in *ACLU*, including its instructions to the trial court to better consider proposals to de-identify the data, strongly supported the conclusion that anonymizing data does not constitute the creation of a record under the CPRA. (*See Exide*, slip. op. at 20–21.)

Other cases have similarly held that agencies may be required to manipulate existing records and databases to extract information sought by a public records requestor and that doing so does not create a new record. In *CBS Broadcasting Inc. v. Superior Court* (2001) 91 Cal.App.4th 892, the Court of Appeal ordered the California Department of Social Services to go through its records and compile and produce accurate lists of individuals granted a criminal conviction exemption to work in licensed child day care facilities and the identity of each facility employing such individuals.

In *May v. Department of Air Force* (5th Cir. 1986) 800 F.2d 1402, a federal appellate court addressed records maintained as handwritten forms that, if released, could reveal the identity of the author based on the distinctive style of the handwriting. The court held the Air Force could create a typewritten copy of the records or recreate them in a third-party's hand to protect the identity of the author. The court held, "such disclosure would . . . ensure maximum disclosure under the [FOIA], and not unreasonably burden the agency." (*Id.* at 1403.)

In *Schladetsch v. H.U.D.* (D.D.C., Apr. 4, 2000, No. 99-0175) 2000 WL 33372125, the court held that neither the programming necessary to instruct a computer to conduct a search, nor the process of extracting and compiling the data resulting from such a search, constitute the creation of a new record. (*Id.* at *3.)

And in *Disabled Officer's Association v. Rumsfeld* (D.D.C. 1977) 428 F.Supp. 454, the court held that the fact that the agencies "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 it preclude the applicability" of FOIA. (*Id.* at p. 456.)

The protocols proposed by Petitioners are more appropriately compared to requirements to redact exempted information from otherwise non-exempt records, which are well-established under both the CPRA and FOIA. (*Compare* § 6253(a); 5 U.S.C. § 552(a)(8)(A).) Redaction within records promotes the goals of the CPRA because it allows for the maximum disclosure of information while still protecting other interests. For example, in *CBS, Inc. v. Block*, the court held that releasing applications for concealed weapon licenses but deleting certain confidential information from those applications protected the privacy of applicants while still ensuring the public was provided with enough information to determine whether public officials were acting properly in issuing licenses for legitimate reasons. (*CBS, Inc. v. Block*, (1986) 42 Cal.3d 646, 655.)

Techniques for protecting exempt information while still releasing otherwise non-exempt government records that are of great interest to the public must evolve as the government's means of collecting, compiling, and maintaining such records has evolved. Protocols that propose to anonymize data, such as those presented by Petitioners, represent one such technique. California courts should not avoid a determination of whether anonymization can protect privacy by dismissing it out of hand as the creation of a "new record."

Conclusion

The Court of Appeal's "new record" holding compromises California's fundamental commitment to transparency, accountability, and access to public records. Amicus respectfully requests that this Court grant the Petition for Review to ensure the public can continue to access government records in the digital age.

Respectfully submitted,

Aaron Mackey
Electronic Frontier Foundation
SBN 286647

Jennifer Lynch
Electronic Frontier Foundation
SBN 240701

SANDER v. S.C. (STATE BAR OF CALIFORNIA)
Supreme Court of the State of California Case No. S251671

CERTIFICATE OF SERVICE

STATE OF CALIFORNIA, COUNTY OF SAN FRANCISCO

I am over the age of 18 years and not a party to the within action. My business address is 815 Eddy Street, San Francisco, California 94109.

On November 1, 2018, I served the foregoing document entitled:

***AMICUS CURIAE* LETTER OF ELECTRONIC FRONTIER FOUNDATION IN SUPPORT OF PETITION FOR REVIEW**

on the attached Service List

X BY ELECTRONIC TRANMISTION VIA TRUEFILING: I caused a copy of the foregoing documents to be sent via TrueFiling to the persons at the e-mail addresses listed in the Service List. The following parties and/or counsel of record are designated for electronic service in this matter on the TrueFiling website. I did not receive, within a reasonable time after the transmission, any electronic message or other indication that the transmission was unsuccessful.

X BY FIRST CLASS MAIL: I caused to be placed the envelope for collection and mailing following our ordinary business practices. I am readily familiar with this firm's practice for collecting and processing correspondence for mailing. On the same day that correspondence is placed for collection and mailing, it is deposited in the ordinary course of business with the United States Postal Service, in a sealed envelope with postage fully prepaid.

I declare under penalty of perjury under the laws of the State of California that the foregoing is true and correct.

Executed on November 1, 2018 at San Francisco, California.

Madeleine Mulkern

SERVICE LIST

Via E-File Service

Jean-Paul Jassy
Kevin Lester Vick
Jassy Vick Carolan LLP
800 Wilshire Boulevard, Suite 800
Los Angeles, CA 90017

Attorneys for Richard Sander:
Petitioner

James M. Chadwick *Via E-File Service*
Sheppard, Mullin, Richter & Hampton
Four Embarcadero Center, 17th Floor
San Francisco, CA 94111-4109

Andrea Nicole Feathers
Sheppard Mullin Richter & Hampton LLP
333 S Hope Street 43rd Floor
Los Angeles, CA 90071

Guylyn Remmenga Cummins
Sheppard Mullin Richter & Hampton LLP
501 W Broadway, 19th Floor
San Diego, CA 92101

David Edward Snyder
California First Amendment Coalition
534 4th Street, Suite B
San Rafael, CA 94901

Attorneys for First Amendment Coalition:
Petitioner

James M. Wagstaffe *Via E-File Service*
Michael John Von Loewenfeldt
Melissa Perry
Kerr & Wagstaffe LLP
101 Mission Street, 18th Floor
San Francisco, CA 94105

Vanessa Lynne Holton
Destie Lee Overpeck
Office of General Counsel
State Bar of California

180 Howard Street
San Francisco, CA 94105

Attorneys for State Bar of California: Real Party in Interest

James M. Wagstaffe
Michael John Von Loewenfeldt
Kerr & Wagstaffe LLP
101 Mission Street, 18th Floor
San Francisco, CA 94105

Via E-File Service

Attorneys for Board of Trustees of the State Bar: Real Party in Interest

San Francisco Superior Court
Civil Center Courthouse
400 McAllister Street
San Francisco, CA 94102

Via First Class Mail