Filed 8/23/18

CERTIFIED FOR PUBLICATION

IN THE COURT OF APPEAL OF THE STATE OF CALIFORNIA

FIRST APPELLATE DISTRICT

DIVISION THREE

RICHARD SANDER et al., Petitioners and Appellants, v. STATE BAR OF CALIFORNIA et al., Defendants and Respondents.	A150061, A150625 (City & County of San Francisco Super. Ct. No. CPF08508880)

Appellants and petitioners Richard Sander and the First Amendment Coalition (Petitioners) challenge the trial court's denial of their petition for writ of mandate seeking to obtain information from the State Bar of California's bar admissions database. Specifically, Petitioners seek individually unidentifiable records for all applicants to the California Bar Examination from 1972 to 2008 in the following categories: race or ethnicity, law school, transfer status, year of law school graduation, law school and undergraduate GPA, LSAT scores, and performance on the bar examination. Making these records available to the public in a manner that protects the applicants' privacy and anonymity, they believe, will allow researchers to study the potential relationship between preferential admissions programs in higher education and a gap in bar passage rates between racial and ethnic groups.

Following a bench trial, the superior court upheld the State Bar's denial of Petitioners' request on five independent grounds. We need address only the first of them. The court correctly found Petitioners' request to be beyond the purview of the California

Public Records Act (Gov. Code § 6250 et seq. (CPRA)) because it would compel the State Bar to create new records.[1] Accordingly, we affirm.

BACKGROUND

I. Phase I Litigation

This case has taken a long and well-documented path between the trial court, this court and the Supreme Court. *Sander v. State Bar of California* (2013) 58 Cal.4th 300, 307–309 (*Sander*) relates much of the relevant background (incorporated by reference here) and provides the starting point for this appeal, the second in this court. (See *Sander v. State Bar of California* (2011) 196 Cal.App.4th 614 [126 Cal.Rptr.3d 330], review granted Aug. 25, 2011 and superseded by *Sander*)

At issue in *Sander* was the trial court's determination that the State Bar has no common law or state Constitutional duty to disclose records in its admissions database. The Supreme Court reversed and remanded. It explained: "The question presented is whether any law requires disclosure of the State Bar's admissions database on bar applicants. We conclude that under the common law right of public access, there is sufficient public interest in the information contained in the admissions database such that the State Bar is required to provide access to it if the information can be provided in a form that protects the privacy of applicants and if no countervailing interest outweighs the public's interest in disclosure. Because the trial court concluded that there was no legal basis for requiring disclosure of the admissions database, the parties did not litigate, and the trial court did not decide, whether and how the admissions database might be redacted or otherwise modified to protect applicants' privacy and whether any countervailing interests weigh in favor of nondisclosure. Consequently, the Court of appeal will be directed to remand the case to the trial court." (*Sander, supra*, 58 Cal.4th at p. 304; see p. 309 ["there is a public interest in access to the State Bar's admissions database that will require the State Bar to disclose the requested information if it can be

[1] Unless otherwise indicated, further statutory citations are to the Government Code.

applied in a form that does not violate the privacy of applicants and if other considerations do not warrant nondisclosure"]; see also *Sander v. State Bar of California, supra,* 196 Cal.App.4th 614.)

The Supreme Court in *Sander* expressly left open whether changes to the admissions database necessary to protect applicants' privacy would entail the creation of new records, and thereby exceed the scope of disclosure required under public access laws. (*Sander, supra*, 58 Cal.4th at p. 327.) It observed that "in the context of electronic records, and in particular electronic databases, to resolve this issue would require consideration of the complexity of the tasks required to produce the data in the form requested; consequently, it would be premature for us to attempt to resolve this issue." (*Ibid.*) The case was remanded for the trial court to resolve (1) whether the requested information can be provided in a form that protects the privacy of applicants; and (2) whether countervailing interests outweigh the public's interest in disclosure. (*Ibid.*)

Following remand, more than a dozen individuals who had applied to take the California Bar Exam since 1972 and two non-profit professional associations of African American lawyers, the Black Women Lawyers Association of Los Angeles, Inc. and the John M. Langston Bar Association of Los Angeles, intervened on the side of the State Bar. They did so "to protect privacy and reputational interests that are at the heart of the litigation between Petitioners and the State Bar" by preventing disclosure of "the sensitive, private information of Individual Intervenors or the members of Organizational Intervenors that stands to be exposed if the Petitioners prevail in this action."

II. The Admissions Database

Subject to a stringent stipulated protective order, the State Bar provided Petitioners' experts with highly confidential data from its admissions databases to facilitate expert analysis concerning the issues to be tried upon remand. The admissions databases consist of five separate text files containing for each applicant who sat for the bar examination between 1972 and 2008: (1) the number of times the applicant took the exam, and whether he or she passed or failed; (2) law school graduation date and a code for any law

school attended; (3) LSAT score and law school GPA; (4) race and ethnicity; (5) a file linking the law school codes and law school names.

III. Post-*Sander* Legislation

In 2015 the Legislature enacted Business and Professions Code section 6026.11, which for the first time made records of the State Bar subject to production under the CPRA. (Stats. 2015, ch. 537, § 6.) It also enacted a specific confidentiality statute governing State Bar admission records. With exceptions not relevant here, section 6060.25 provides that "Notwithstanding any other law, any identifying information submitted by an applicant to the State Bar for admission and a license to practice law and all State Bar admission records, including, but not limited to, bar examination scores, law school grade point average (GPA), undergraduate GPA, Law School Admission Test scores, race or ethnicity, and any information contained within the State Bar Admissions database or any file or other data created by the State Bar with information submitted by the applicant that may identify an individual applicant . . . shall be confidential and shall not be disclosed pursuant to any state law, including, but not limited to, the California Public Records Act. . . ." (Bus. & Prof. Code, § 6060.25, subd. (a), added by Stats 2015, ch. 537, § 8.)

IV. The "Phase II" Trial

The "Phase II" issues were tried to the court over five days. Much of the trial was devoted to competing expert testimony about whether disclosure of the admission records may reveal bar applicants' private information or require the State Bar to create new records not already in its possession.[2] A primary issue was whether the four different protocols proposed by Petitioners to de-identify or "anonymize" the data were sufficient

[2] Four of the Intervenors testified to their concerns about the release of information they provided to the State Bar with the understanding it would remain confidential. State Bar admissions director Gayle Murphy testified, *inter alia,* about the Bar's collection and treatment of confidential information from applicants for admission to the bar and the State Bar's response to Petitioners' requests.

to prevent matching a record in the supposedly anonymous data to either an individual or a small group of individuals.

A. Data Anonymization and Re-Identification

Dr. Latanya Sweeney, a leading expert in data privacy and anonymization, provided a 67-page report and testified for the State Bar about the risks to bar candidates' privacy in releasing data in the manner proposed by Petitioners. Dr. Sweeney is a professor at Harvard University, where she leads the Data Privacy Lab and teaches classes in data privacy. She holds a Master's Degree in computer science and electrical engineering and a Ph.D. in computer science from M.I.T. Her work in the field of data privacy has been cited in over 5,000 scientific publications.

Dr. Sweeney testified that re-identification refers to whether someone can "use reasonable effort to match the person's identity to details in the released dataset sufficient to know enough information about the person to identify him or her as a specific person." "We use the term 'named person' to refer to having sufficient information to individually identify the person who is the subject of the data. Thus, if records in the DataSet can be associated with named people, then the DataSet would be re-identified. Harm from a re-identification can result if sensitive information contained in the data becomes known about named persons. Although in most circumstances privacy concerns relate to identification of an individual by strangers, in some cases, targeted identification of a specific known person, and even self-identification, can be problematic where either the data is not intended to be known to the person to whom it pertains or the facts that enable self-identification can be compelled (such as by a prospective employer) even where they are not generally known to the public. In each case the question is the same: can the 'anonymous' data be re-identified such that information is learned about specific individuals?"

Dr. Sweeney explained: "A unique re-identification occurs when a record in the data matches to one person's information uniquely. A 'group re-identification' occurs when a few records in the dataset match to a small number of people. Both are examples of re-identification that raise privacy concerns. For example, if a proposed data release does

not include names or home addresses, but did include exact birthdates and specific employers, it would be possible to use publicly-available websites and directories to deduce the identity of many individuals in the database. A few-to-few match can be just as damaging as a one-to-one match. For example, a release of medical records showing that six of seven named individuals have a genetic disposition toward cancer would result in a group re-identification of each individual as likely (6 in 7) to have that condition even though it was not known which of the individuals was the one without the condition. It is well recognized that this type of few-to-few identification poses privacy risks similar to unique identification." Data analytic companies that deal in data compilations and derivations to link disparate datasets are also becoming increasingly able to re-identify data historically regarded as anonymous.

B. Petitioners' De-Identification Protocols

Petitioners' experts proposed four methods or "protocols" for rendering data anonymous that they believed could protect privacy rights without unduly burdening the State Bar. Under Protocol One, the State Bar would set up a physical "data enclave" to house a version of the admissions database stripped of personal identifiers (name, address, social security numbers and the like) and specified records (e.g., records of students who attended unaccredited or correspondence schools or whose race is coded American Indian, Alaska native, Filipino, Pacific Islander or from the Indian subcontinent and for law schools that graduated fewer than 10 students who took the bar exam in a given year). These processes would exclude approximately 30 percent of all student records. The remaining data would be converted into a format compatible with a statistical analysis software package and maintained in a "safe room" where members of the public could conduct research under the supervision of an on-site operator. Anyone seeking access to the data would have to explain their purpose for seeking the information and sign an agreement not to re-identify individual applicants from the data. Once granted access, users could use only electronic equipment and software provided within the safe room and would be strictly limited as to hours of access and the kind of information they could take away.

Under Protocols Two, Three and Four, petitioners proposed to apply various techniques to the admissions database, such as data redaction, recoding and binning, that would conceal applicant identities and prevent the risk of reidentification.[3] Protocols 2 and 4 employ variations of a concept known as "k-anonymity," a leading method of anonymizing data invented by Dr. Sweeney. She explained, "The concept is simple. A dataset is k-anonymous if for every record that appears in the data there are k indistinguishable copies, where k is greater than or equal to 2. The scientific guarantee flows directly from the definition. If a record contains information about a person, then any record will map to at least k individuals; and conversely, if you know a person whose information is in the data, then there will be at least k records that could belong to the person. A k-anonymized dataset also maintains its privacy guarantee if the dataset is linked to other data, because any attempt to link to a record in the data will do so ambiguously to at least k records. Any re-identification attempt can never do better than k equally likely matches. The proper size of k (i.e. how much anonymity to provide) is a policy question depending on the sensitivity of the data and how ambiguous an identification is considered permissible." In other words, "K-anonymity, its privacy guarantee is that for every record I see in the data, there are going to be at least k other – there are going to be k copies of that record that are indistinguishable."

Applying k-anonymity to a dataset requires both choosing the size of k and the data fields (e.g., bar passage, school, graduation year and race), also called variables or attributes, to be anonymized.[4] After direct identifiers such as names and social security numbers are removed and the data is cleaned, various techniques can be used to anonymize the remaining data. These include suppression (removing information from

[3] Binning refers to the practice of grouping and segregating data of reasonably equivalent values into a single group or set.

[4] A particular instance of the variable is called a "value." A row of data is an "observation." For any particular set of variables, a group of observations that share the same value for those variables is a "cell."

data that might be identifying), adding "random noise," scrambling data or generalizing fields of information, or swapping values for generalized values.

Under Protocols 2 and 4 Petitioners would apply a k of 11 to four different variables: law school, graduation year, race, and whether the person ever passed the bar.[5] This would be achieved by eliminating "substantial portions" of the admissions database, reducing the eight ethnic categories in the data to four (white, black, Hispanic, and other); collapsing graduation year data for law schools into three, six or nine-year "bands," depending on the average number of the school's graduating class who took the bar, and for the smallest schools conducting various data manipulations to mask information.[6] If those steps produced any groups (or k's) of fewer than 11 identical records, the race variables would be merged into aggregate categories for either " 'underrepresented minority' " or " 'White and Other' " or, if either category still yielded groups with fewer than 11 members, race would be redacted.

Protocol 3 did not employ K-anonymity. Instead, it principally relied upon a statistical analysis or, "more precisely, it describes intense computations of mathematically determined standardizations of the data that report relative positions within various distributions found in the data." According to Dr. Sweeney, "it is clear that this protocol requires a great deal of effort, even for me writing a Python [software] program." "The first step in Protocol 3 removes huge portions of the data. All data from 1982-1984, and 1999-2005 are removed. Records from 1985-1998 and 2006-2008 remain. The second step recodes race into the same four categories used in Protocol 2 and Protocol 4—namely, black, white, Hispanic and other. Then come the steps involving statistical computations. The computations are done only on those records for

[5] They would not k-anonymize LSAT or GPA scores or whether the applicant took the bar more than once. Dr. Sweeney opined that these data fields are "potentially knowable" by third parties and, if not anonymized, can be used to re-identify applicants from the dataset using personally known or searchable public information.

[6] Protocol 1 would also employ k-anonymity, but using a k of 5.

law schools having 20 or more LSAT scores in a given year; all other records are dropped. The computations themselves require computing different averages and standard deviations and then recoding the data with new LSAT fields that contain the relative position of the original LSAT score within the distributions computed." The next step in Protocol 3 manipulates the GPA field similarly, while the year of graduation is replaced with a 4-year period and, finally, law school names are (with some exceptions) replaced with "California" or "out of state." Petitioners' experts describe Protocol 3 as "a more radical approach to de-identifying the data" and conceded it was "not a method we recommend, because of its impact upon data utility."

Protocol Four, as described by Petitioners' experts, incorporated Protocol Two, and (1) randomly redacted the applicant's law school for 25% of the observations (see *infra,* fn. 4); (2) rounded or suppressed law school GPA to no more than 2 digits; and (3) redacted unique GPA values.

Dr. Sweeney concluded that "[a]lthough the level of risk varies amongst the protocols, each of Petitioners' proposals for releasing the State Bar admissions data presents cognizable risks that individuals may be specifically identified in the data, and thus their bar scores, academic history, and other private information publicly revealed. Not only *may* the information reveal specific individuals, as I have demonstrated it clearly *does* reveal information on specific individuals. [¶] I am not opining that it is impossible to anonymize the data, quite the contrary. However, the proper way to anonymize this kind of sensitive data requires anonymization of all fields in the data . . . and to do so using scientifically proven methods, not ad hoc binning, including replacing substantial portions of the data with more generalized data or codes, and potentially adding additional fictitious data. Petitioners' protocols come nowhere close to meeting those standards or otherwise assuring that none of the individuals in the DataSet can be identified with reasonable certainty to their data, and further, some of the protocols have a disproportionate adverse impact on underrepresented minorities."

Admissions director Murphy testified that the State Bar does not maintain admissions data in the clustered and banded formats Petitioners were requesting and that it would have to create new documents to provide the requested records.

Petitioners introduced an expert report that explained their protocols and disagreed with Dr. Sweeney's analysis. Their expert, Luk Arbuckle, testified about the efficacy of the protocols to produce useful data while protecting the privacy of bar applicants.[7] In contrast to Dr. Sweeney's view, Arbuckle opined that the proposed protocols adequately mitigated the risk that individual applicants could be identified. "All of the Protocols we have described take the most important privacy step: they all eliminate personal information from the Admissions Database, including names, addresses, and Social Security numbers. They then each take one or more supplemental steps to prevent any data user from inferring individual identity by comparing information in the disclosed data with outside information sources. In Protocol One, this is done by maintaining the data in an Enclave, which prevents users from bringing in outside data and thus making comparisons of 'inside' and 'outside' information. In Protocols Two and Four, the key supplemental steps are to make sure that 'cells' – i.e., combinations of information about bar applicants that can be generally uncovered from other public sources, with sufficient research – contain no fewer than eleven observations. This makes it impossible to be certain of any *individual's* identity, using only the specified 'publicly knowable' variables. Protocol Four takes additional protective measures to guard against those rare cases where a bar-taker['s] Law School GPA might be known or inferred from public sources" and "uses random censoring of a sample of law school identities. . . . In Protocol Three, two supplemental procedures are used: recoding individual LSAT scores

[7] Arbuckle holds Master of Science degrees in in Statistics and Mathematics. He had five years' experience in the field of data anonymization and re-identification risk at the time of trial. In addition, labor economist Dr. Peter Arcidiacono testified for Petitioners about the usefulness of the de-identified datasets under the different protocols for economic and social research. Petitioners also presented deposition testimony from Dr. Felicia LeClere, who developed and analyzed the protocols with Arbuckle, and social science researcher Samuel Canas.

and law school GPAs as standardized values, and then deleting law school identifiers from the database." Arbuckle disagreed with Sweeney's opinion that the failure to K-anonymize *all* data fields introduced an unacceptable risk of re-identification using known or searchable information, principally because, in Petitioners' view, such information was not publicly available.

The court denied the petition in a detailed 22-page order on five independent grounds. "1. The disclosure of the requested records pursuant to any of Petitioners' protocols requires the creation of a new record, and thus the State Bar is not required to disclose the data; [¶] 2. The requested records are barred from disclosure pursuant to Business and Professions Code section 6060.25; [¶] 3.The requested records are exempt from disclosure pursuant to Government Code section 6254(k) in that disclosure is prohibited by state law; [¶] 4. The requested records are exempt from disclosure pursuant to Government Code section 6254(c) in that disclosure constitutes an unwarranted invasion of privacy; and [¶] 5. The requested records are exempt from disclosure pursuant to Government Code section 62255(a), the CPRA's "catch-all" exemption." Judgment was entered for the State Bar and the intervenors and against Petitioners.

Petitioners filed a timely appeal and petition for writ of mandate from the superior court judgment. We consolidated the appeal and the writ petition and issued an order to show cause why the petition should not be granted. We have also considered amicus curiae briefs filed on Petitioners' behalf by the Reporters Committee for Freedom of the Press and 13 media organizations; the Pacific Legal Foundation; the National Association of Scholars; Gail Heriot and Peter Kirsanow; and the Electronic Frontier Foundation.

DISCUSSION

The court's initial ground for denying the petition is that disclosure of the bar admissions data would require the Bar to create new records, a duty not imposed by California's access to public records laws. Petitioners contend this was erroneous as a matter of law and that undisputed evidence shows the data manipulations required to institute their proposed protocols are within the duties imposed by the CPRA. We disagree.

I. Standard of Review

We independently review the trial court's interpretation of the CPRA and its application to undisputed facts, but accept the court's findings of historical fact if supported by substantial evidence. (*American Civil Liberties Union of Northern Cal. v. Superior Court* (2011) 202 Cal.App.4th 55, 66; *BRV, Inc v. Superior Court* (2006) 143 Cal.App.4th 742, 750; *Fredericks v. Superior Court* (2015) 233 Cal.App.4th 209, 223–224 (*Fredericks*); *Regents of University of California v. Superior Court* (2013) 222 Cal.App.4th 383, 397 (*Regents*).)

" ' "A court's overriding purpose in construing a statute is to ascertain legislative intent. . . . [Citation.] In interpreting a statute to determine legislative intent, a court looks first to the words of the statute and gives them their usual and ordinary meaning. [Citation.] Statutes must be given a fair and reasonable interpretation, with due regard to the language used and the purpose sought to be accomplished." ' " (*Fredericks, supra*, 233 Cal.App.4th at p. 224.) Moreover, "[i]n CPRA cases, this standard approach to statutory interpretation is augmented by a constitutional imperative. [Citation.] Proposition 59 amended the Constitution to provide: '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.' " (*City of San Jose v. Superior Court* (2017) 2 Cal.5th 608, 617; Cal. Const., art I, § 3, subd. (b).)

II. Analysis

"The core purposes of the CPRA are to prevent secrecy in government and to contribute significantly to the public understanding of government activities." (*Fredericks, supra*, 233 Cal.App.4th at p. 223.) To that end, the CPRA directs that "[e]xcept with respect to public records exempt from disclosure by express provisions of law, each state or local agency, upon a request for a copy of records that reasonably describes an identifiable record or records, shall make the records promptly available to any person upon payment of fees covering direct costs of duplication, or a statutory fee if applicable. Upon request, an exact copy shall be provided unless impracticable to do so."

(§ 6253, subd. (b); see also § 6253.9, subd (a) [CPRA applies to records maintained in electronic format].)

As manifested by this case, an unavoidable tension exists between the CPRA's laudable purposes of transparency and disclosure and " 'the equally important public interest in protecting citizens and public servants from unwarranted exposure of private matters.' " (*Fredericks, supra*, 233 Cal.App.4th at p. 223.) To accommodate those often-competing interests, "[t]he CPRA generally presumes that all documents maintained by a public entity are subject to disclosure to any member of the public, *unless a statutory exemption applies or the catchall exemption, section 6255, is satisfied* (when public interest served by nondisclosure of records clearly outweighs the public interest in disclosure)." (*Ibid*, italics added; see §§ 6254 et seq. [exemptions to disclosure].)

The threshold question in this case is not whether the information Petitioners seek is subject to one of the CPRA's statutory exemptions from disclosure. Nor is it whether, as Petitioners would have it, the trial court improperly created a nonstatutory "new exemption" for the records sought. The question, rather, is whether the information in the form Petitioners ask the State Bar to release it is subject to the obligations imposed by the CPRA in the first instance. The trial court correctly concluded that it is not.

Petitioners' threshold claim is that the trial court got the burden of proof backward by implicitly "requiring them to prove that specific de-identification protocols would prevent any individual from being re-identified and would prevent any group of individuals [from] having any information revealed about it," rather than requiring the State Bar to prove the data could not be produced in a way that would adequately protect bar applicants' privacy. (See *International Federation of Professional and Technical Engineers, Local 21, AFL-CIO v. Superior* Court (2007) 42 Cal.4th 319, 329 ["The party seeking to withhold public records bears the burden of demonstrating that an exception applies"].) We do not need to separately address this argument. The trial court's finding that the protocols require the creation of new records is independent of its conclusion on

the possibility of re-identification. Petitioners' argument has no bearing on our analysis or holding on whether the protocols require creation of new records.

In any event, the argument is not convincing. The trial court expressly stated that "the State Bar has demonstrated that disclosure of the requested records is prohibited by Business and Professions Code 6060.25 because individual applicants may be identified from the data resulting from application of any of Petitioners' protocols. *Accordingly, the Court finds that the* S*tate Bar has met its burden.*"[8] (Italics added.) The trial court also expressly acknowledged the State Bar's burden under section 6255, subdivision (a) (the "catch-all" exemption) to show that the public interest served by nondisclosure outweighed the public interest served by disclosure: "The CPRA provides that a public agency is justified in withholding records *if it demonstrates* that 'the public interest served by not disclosing the record clearly outweighs the public interest served by disclosure of the record.' Gov't Code § 6255(a)." (Italics added.) We will not infer error where it is not shown on the record.

But, as we have indicated, the trial court was correct for another, independent reason. It is well established under California law and guiding federal precedent under the Freedom of Information Act (FOIA) (see *Regents*, *supra*, 222 Cal.App.4th at p. 400) that, while the CPRA requires public agencies to provide access to their existing records, it does not require them to create new records to satisfy a request. (*Fredericks, supra,* 233 Cal.App.4th at p. 227]; *Haynie v. Superior Court* (2001) 26 Cal.4th 1061, 1073–1075 [agency not under duty to create a log of potentially responsive records].) "The basic rule

[8] Business and Professions Code section 6060.25, subdivision (a) provides: "Notwithstanding any other law, any identifying information submitted by an applicant to the State Bar for admission and a license to practice law and all State Bar admission records, including, but not limited to, bar examination scores, law school grade point average (GPA), undergraduate GPA, Law School Admission Test scores, race or ethnicity, and any information contained within the State Bar Admissions database or any file or other data created by the State Bar with information submitted by the applicant that may identify an individual applicant, other than information described in subdivision (b), shall be confidential and shall not be disclosed pursuant to any state law, including, but not limited to, the California Public Records Act."

is that an agency must comply with a request if responsive records can be located with reasonable effort. [Citation.] If the agency would be required to create a new set of public records in order to provide responses to a CPRA request, such agency action may be found to exceed its statutory duties." (*Fredericks, supra,* 233 Cal.App.4th at p. 227; see *Regents, supra,* 222 Cal.App.4th at p. 400 ["Similarly to the FOIA, no language in the CPRA creates an obligation to create or obtain a particular record when the document is not prepared, owned, used or retained by the public agency"].)

Federal law construing the FOIA is in accord. "The Act does not obligate agencies to create or retain documents; it only obligates them to provide access to those which it in fact has created and retained. . . . [O]nly the Federal Records Act, and not the FOIA, requires an agency to actually create records, even though the agency's failure to do so deprives the public of information which might have otherwise been available to it." (*Kissinger v. Reporters Committee for Freedom of the Press* (1980) 445 U.S. 136, 152, citing *NLRB v. Sears, Roebuck & Co.* (1975) 421 U.S. 132, 161–162 [agency had no duty to create material explaining information in disclosed records; FOIA "only requires disclosure of certain documents which the law requires the agency to prepare or which the agency has decided for its own reasons to create"]; *Yeager v. Drug Enforcement Administration* (D.C. Cir. 1982) 678 F.2d 315, 321 (*Yeager*); *Students against Genocide v. Department of State* (D.C. Cir. 2001) 257 F.3d 828, 837 ["although agencies are required to provide 'any reasonably segregable' non-exempt portion of an existing record, they are not required to create new documents"]; *Center for Public Integrity v. F.C.C.* (2007) 505 F.Supp.2d 106, 114 (*Center for Public Integrity*).)

We have found no cases addressing proposals for data manipulations as complex as those proposed by Petitioners, but *Center for Public Integrity, supra,* 505 F.Supp.2d 106, is instructive. The plaintiff there initially requested records submitted by telecommunications providers to the FCC but withdrew that request during litigation and proposed instead that the FCC "replace filers' numerical responses with either ranges . . . or an indication of whether the deleted responses were 'zero or greater than zero.' " (*Id.* at p. 114.) The District Court for the District of Columbia ruled that requiring the FCC to

produce data in either of those forms would require the creation of new records. "[P]laintiff's proposal would require the FCC to do more than simply redact portions of the numbers . . . . The FCC would also have to replace the redacted numbers with new numbers, which the FCC itself would have to select. *Because agencies are not required to create new records to satisfy FOIA requests, the Court is without authority to require the FCC to adopt plaintiff's proposal for the disclosure of modified data. . . ."* (*Ibid*, italics added, citing *Students Against Genocide v. Department of State, supra,* 257 F.3d 828 at p. 837 [agency could not be ordered to produce altered photographs at different resolutions from the originals to mask the capabilities of the systems that took them]; see *American Civil Liberties Union v. Arizona Department of Child Safety* (Ariz. App. 2016) 377 P.3d 339, 345–346 [construed consistently with FOIA, Arizona public records law requires agency to search its electronic database to produce existing records but does not require it "to create a new record that compiles analytical information about information"].)

Petitioners argue that much of this authority predates the emergence of electronic databases as a commonplace repository of government information, and that more recent cases require disclosure of electronically stored information "even if it requires extensive compilation or extraction of data contained in electronic public records." Their argument is premised upon mischaracterizing the cited cases, which merely distinguish between searching, extracting, compiling or redacting electronically stored data, which our state and federal public access laws require, and creating new records, which they do not. (See *Schladetsch v. U.S. Dept. of H.U.D* (D.D.C. 2000) 2000 WL 33372125 [programming necessary to perform computer search to extract and compile data did not amount to creating a new record]; *International Diatomite Producers Ass'n. v. U.S. Social Security Admin.* (N.D. Cal 1993) 1993 WL 137286 [redaction and segregation of data is not equivalent to creating a new record]; *Osborn v. Board of Regents of University of Wisconsin System* (Wis. 2002) 254 Wis. 2d 266, 299–302 [extraction and compilation to segregate exempt from non-exempt data]; *Bowie v. Evanston Community Consol.*

School Dist. No. (Ill. 1989) 128 Ill.2d 373, 376, 382 [holding that disclosing student test scores in a "masked and scrambled format" did not create a new record].)

Here, the trial court determined that each of Petitioners' four proposed protocols would require the creation of new records. "All of the protocols require the State Bar to recode its original data into new values. . . . For example, the protocols group law schools into three classes, designating a 'school class' code, which is not present in the original Admissions Database. [Citations.] The protocols also involve recoding race/ethnicity values to reflect four categories (Asian, Hispanic, Black, or White) instead of the State Bar's original eight race categories. Similar codes are created with respect to year of graduation. [Citations.]

"In addition, Protocol One would also require the State Bar to create a physical data enclave which would provide restricted access to the State Bar's Admissions Database. Petitioners, however, failed to present any authority in support of their position that the CPRA allows this Court to order the State Bar to create this type of data enclave. The type of data enclave proposed under Protocol One is simply not a valid remedy under the CPRA.

"Protocols Two, Three, and Four require the creation of even more new data. For example, Protocol Two involves replacing some applicants' actual LSAT scores with a calculated median, as well as possibly creating a new 'underrepresented minority' or 'URM' category that does not exist in the original Admissions Database. [Citation.] Protocol Four involves rounding off actual law school GPAs to two significant digits. [Citation.] Finally, Protocol Three, which both the State Bar's and Petitioners' experts agree requires drastic changes to the State Bar's original data, requires calculating new values for GPAs and LSAT scores, as well as creating a variable indicating whether an applicant's law school is located in California or out-of-state. [Citations.]" The court cited Dr. Sweeney's testimony that none of the variables in Protocol Three exist in the raw Bar data, and that every variable would have to be calculated or recoded or both, as well as testimony from one of Petitioners' experts that " '[s]o much information has been

changed or removed entirely from the data, law school name, the exact GPA, LSAT, all the data cleaning steps, they do a lot to change the structure of the data.' "

The court also rejected the Petitioners' contention that their protocols merely required the State Bar to redact or manipulate existing data and do some computer programming. "It is clear that the various steps outlined do more than simply redact or omit existing data. In order to achieve the 'manipulated' data contemplated under each of the protocols, Petitioners had to produce a 'Stata' software that applied a code specifically created to generate new data. [Citations.] Indeed, this case is vastly distinct from the two Illinois district court cases cited by Petitioners, in which the public agencies were ordered to produce a computer program that could *delete* certain information." Requiring the Bar to recode its existing data, the court concluded, would thus require it to create new records. We agree.

Petitioners argue that two provisions of the CPRA demonstrate that it in fact does impose a duty on public agencies to create new records. Section 6253.9, subdivision (b) authorizes an agency to charge a party requesting electronic records "the cost to construct a record, and the cost of programming and computer services necessary to produce a copy of the record" if the request "would require data compilation, extraction, or programming to produce the record." Section 6253, subdivision (c)(4) permits an agency to delay its response to a request based on the time it takes to "compile data, to write programming language or a computer program or to construct a computer report to extract data." Neither requires reprogramming computerized data to create new records. No one disputes that public agencies can be required to gather and segregate disclosable electronic data from nondisclosable exempt information, and to that end perform data compilation, extraction or computer programming if "necessary to produce a copy of the record." (§ 6253.9, subd. (b).) But segregating and extracting data is a far cry from requiring public agencies to undertake the extensive "manipulation or restructuring of the substantive content of a record" (*Yeager, supra,* 678 F.2d at p. 323) such as Petitioners propose here. Certainly, they have not identified any instances in which courts have compelled a public agency to undertake programming that would assign new or different

values to existing data, replace groups of data with median figures or variables, and collapse and band data into newly defined categories.

In short, the trial court got the law right. There is no doubt that a government agency is required to produce non-exempt responsive computer records in the same manner as paper records and can be required to compile, redact or omit information from an electronic record. (See, e.g., §§ 6253.9, 6253 subd. (a)); *Sierra Club v. Superior* Court (2013) 57 Cal.4th 157, 165; *Yeager*, *supra*, 678 F.2d at pp. 322-323.) But it cannot be required to create a new record by changing the substantive content of an existing record or replacing existing data with new data. The trial court's application of this distinction was entirely correct.[9]

Petitioners also contend the trial court got it wrong on the facts. In their view, notwithstanding Dr. Sweeney's and Murphy's testimony the evidence established that disclosure pursuant to their protocols "would not create a new record, but would at most require data extraction, compilation and programming." But "[w]hen a trial court's factual determination is attacked on the ground that there is no substantial evidence to sustain it, the power of an appellate court begins and ends with the determination as to whether, on the entire record, there is substantial evidence, contradicted or uncontradicted, which will support the determination, and when two or more inferences can reasonably be deduced from the facts, a reviewing court is without power to substitute its deductions for those of the trial court. If such substantial evidence be found, it is of no consequence that the trial court believing other evidence, or drawing other reasonable inferences, might have reached a contrary conclusion." (*Bowers v. Bernards* (1984) 150 Cal.App.3d 870, 873–874, italics omitted.) The trial court here considered extensive evidence, notably Dr. Sweeney's report and expert testimony. On that evidence, it determined that Petitioners' requests required the creation of new

[9] Our rejection of Petitioner's claim that sections 6253.9 and 6253, subdivision (c)(4) imply a break with settled law that public agencies are not required to create new records also dispenses with Petitioners' argument that federal cases interpreting FOIA on this issue have no bearing on the CPRA because FOIA "has no equivalent provisions."

records, not simply data redaction, deletion or compilation. Because its finding is supported by substantial evidence, we will not disturb it.

Petitioners assert that, if nothing else, the court on its own initiative should have concocted a plan for disclosing the bar application data "subject to a process that entails *only* redaction of information, which would not require creating anything." As we understand it, they premise this suggestion on the requirement that it is the public agency's burden to prove a basis for nondisclosure of a public record. (See § 6255, subd.(a) [agency must show withheld record is exempt from disclosure]; *American Civil Liberties Union of Northern Cal. v. Superior Court* (2011) 202 Cal.App.4th 55, 67, 84–85.) This, they seem to assert, means *the trial court* was obligated to independently formulate a viable plan that would allow the State Bar to provide some, but not all, of the requested fields of data, while protecting bar applicants' privacy interests. It seems odd for Petitioners to expect the trial court to succeed where their own experts in this highly technical field did not. But no matter. Basic rules of appellate procedure prevent us from addressing this claim on its merits. At no time in the trial court proceedings did Petitioners suggest the State Bar was required to provide some, but not all, of the requested fields of data. Nor, as far as we can tell, did they ask the trial court to order the State Bar to do so, or present evidence on which fields could and should be released or how any such release would not jeopardize Bar applicants' privacy. "It is axiomatic that arguments not raised in the trial court are forfeited on appeal." (*Kern County Dept. of Child Support Services v. Camacho* (2012) 209 Cal.App.4th 1028, 1038.)

In summary, the trial court's determination that Petitioners' requests are beyond the purview of the CPRA is legally correct and supported by the record. That finding is an independently sufficient ground to deny the petition, so we need not address the trial court's four additional stated bases for its decision.

DISPOSITION

The judgment is affirmed. The petition for writ of mandate is denied.[10]

[10] We previously deferred ruling on Petitioners' January 17, 2018 request for judicial notice of a 2018 report by the State Bar of California to the Supreme Court titled "Report to the Supreme Court of the State of California: Final Report on the 2017 California Bar Exam Standard Setting Study" and related correspondence. We now deny the request because the documents to be judicially noticed have no bearing on our analysis and disposition here. (*Mangini v. R.J. Reynolds Tobacco Co.* (1994) 7 Cal.4th 1057, 1063, overruled on another point in *In re Tobacco Cases II* (2007) 41 Cal.4th 1257.) " '[J]udicial notice, since it is a substitute for proof [citation], is always confined to those matters which are relevant to the issue at hand.' " (*Ibid*.)

_____
Siggins, P.J.

We concur:

_____
Pollak, J.

_____
Jenkins, J.

Trial Court:	San Francisco City and County Superior Court
Trial Judge:	Honorable Mary E. Wiss

Counsel:

Jassy Vick Carolan, Jean-Paul Jassy, Kevin L. Vick, Duffy Carolan, for Petitioner and Appellant, Richard Sander.

Sheppard, Mullin, Richter & Hampton, James M. Chadwick, Guylyn R. Cummins, Andrea N. Feathers; First Amendment Coalition, David E. Snyder, for Petitioner and Appellant, First Amendment Coalition.

Eugene Volokh, Richard J. Peltz-Steele, Robert E. Steinbuch, counsel for National Association of Scholars, Amicus Curiae in support of Petitioners and Appellants.

Pacific Legal Foundation, Joshua P. Thompson, Timothy R. Snowball, counsel for Pacific Legal Foundation, Amicus Curiae in support of Petitioners and Appellants.

Electronic Frontier Foundation, Jennifer Lynch, Aaron Mackey, Adam Schwartz, counsel for Electronic Frontier Foundation, Amicus Curiae for Petitioners and Appellants.

Reporter's Committee for Freedom of the Press, Katie Townsend, Bruce D. Brown, Michael Shapiro, counsel for Reporter's Committee for Freedom of the Press, American Society of News Editors, Associated Press Media Editors, Association of Alternative Newsmedia, Bay Area News Group, The California News Publishers Association, Californians Aware, The Center for Investigative Reporting, Los Angeles Times Communications LLC, The McClatchy Company, MPA – The Association of Magazine Media, National Press Photographers Association, Online News Association, and Society of Professional Journalists, Amici Curiae for Petitioners and Appellants.

Center for Constitutional Jurisprudence, John C. Eastman, Anthony T. Caso; Gail Heriot, counsel for Peter Kirsanow and Gail Heriot, Amici Curiae for Petitioners and Appellants.

Kerr & Wagstaffe, James M. Wagstaffe, Michael von Loewenfeldt, Melissa Perry; State Bar of California Office of General Counsel, Vanessa Lynne Holton, Destie Lee Overpeck, for Respondents.

Steptoe & Johnson, William F. Abrams, Margaret Pirnie Kammerud, David H. Kwasniewski, for Intervenors and Real Parties in Interest.