BUILDING THE CFPB'S ARBITRATION ARCHIVE: A COMMENTARY ON DESIGN, IMPLEMENTATION, AND PRIVACY

XIAOLING ANG[*]
THOMAS J. KEARNEY[†]

Abstract

The Consumer Financial Protection Bureau is considering a proposal to require the submission of arbitral claims and awards, which could be published online. The proposal's justifications include analyses of trends, diagnostics of arbitrator bias, and transparency in the arbitral system. This article considers the potential role of archive design and collection in the effectiveness of the arbitration archive. We focus on how submission requirements and data structure may affect the usefulness of the archive, especially for empirical analyses. We also discuss the unique privacy issues at stake. The analysis draws on legal considerations applicable to the financial services sector and potentially covered persons, the design of existing government-managed data archives of industry data, and previous studies of arbitral claims and awards.

---

[*] Principal Consultant, Edgeworth Economics. The opinions expressed herein do not necessarily represent those of Edgeworth Economics nor of any other Edgeworth consultant. This article is intended to inform readers about legal developments. Nothing in this article should be construed as legal advice or a legal opinion, and readers should not act upon the information contained in this article without seeking the advice of legal counsel.
[†] Partner, Akerman LLP.

Table of Contents

I. Introduction

An element of the Consumer Financial Protection Bureau's potential rulemaking on arbitration agreements is a proposal to require the submission of arbitral claims and awards.[1] These submissions could be published online by the Consumer Financial Protection

[1] CONSUMER FIN. PROTECTION BUREAU, SMALL BUSINESS ADVISORY REVIEW PANEL FOR POTENTIAL RULEMAKING ON ARBITRATION AGREEMENTS: OUTLINE OF PROPOSALS UNDER CONSIDERATION AND ALTERNATIVES CONSIDERED 3–5 (Oct. 7, 2015), http://files.consumerfinance.gov/f/201510_cfpb_small-business-review-panel-packet-explaining-the-proposal-under-consideration.pdf [https://perma.cc/2SPT-GKJD] [hereinafter OUTLINE OF PROPOSALS] ("To better understand arbitrations that occur now and in the future, the Bureau is therefore considering proposals that would facilitate ongoing Bureau as well as public monitoring of consumer financial arbitrations. Specifically, the Bureau is considering a proposal that would require companies that use arbitration agreements with consumers for certain types of consumer financial products or services to submit claims filed and awards issued in any arbitration proceedings to the Bureau. The Bureau is further considering periodically publishing the claims or awards on its website.").

Bureau (CFPB).[2] The justifications proposed for this approach center around transparency and related diagnostic analyses, and include identifying administrator bias, ongoing monitoring of arbitrations to identify trends in subject matter and outcomes, transparency of awards and arbitration decisions, and instilling confidence in the arbitral system through transparency.[3]

The CFPB's arbitral archive data collection would apply to all individual and class consumer arbitrations[4] related to products in the scope of the CFPB's jurisdiction.[5] The challenges of defining the parameters of the data collection flow not only from the breadth of the products and providers covered,[6] but also from the logistic and privacy considerations of creating a data collection and processing system that is systematic and durable to changes in participants, products, and analysis goals.[7] Assuming that a submission system for consumer financial arbitral claims and awards is put into place, this article considers the costs and benefits of implementing the proposal to require the submission of arbitral claims and awards for consumer financial products,[8] and how these costs and benefits vary with the nature of implementation. The CFPB should take into consideration the nuances and myriad business practices of covered providers when crafting a regulation that touches on all of their related arbitration agreements. Given the varied roles of covered persons and types of consumer interactions with firms in the consumer financial marketplace, the CFPB's task is complex.[9]

The following analysis focuses on implementation considerations and policy implications of the potential arbitral

---

[2] *Id.*

[3] *Id.* at 20.

[4] *See id.* at 21 ("It is important to note that the proposal under consideration would apply equally to individual arbitration proceedings and any arbitration that could proceed on an aggregate basis.").

[5] *Id.* at 12 (referencing the CFPB's authority granted by the Dodd-Frank Act to "prohibit or impose conditions or limitations on the use of an agreement between a covered person and a consumer for a consumer financial product or service providing for arbitration of any future dispute between the parties").

[6] *See id.* at 22–23 (listing five broad categories of covered providers as defined under the Dodd-Frank Act, Section 1002, subject to limitations under Sections 1027 and 1029).

[7] *See id.* at 20.

[8] *See id.* at 19–22.

[9] *See id.* at 23–25.

publication system.[10] Following this Introduction, Part II considers potential applications of the data and its form, both with respect to potential data analyses and privacy considerations that the CFPB should take into account. Part III turns attention to the practicalities of implementation, both in terms of maintaining and distributing data, and in terms of who is responsible for privacy considerations. The approach to data collection, processing, and privacy should be driven by analysis and privacy goals, which should be developed and considered by the CFPB throughout its rulemaking process.

II. Potential Analyses and Privacy Considerations

A. Research and Analysis Applications

Pursuant to the Small Business Regulatory Enforcement Fairness Act (SBREFA), the Small Business Advisory Review Panel gathered on October 7, 2015 to discuss the potential rule on arbitration agreements.[11] The CFPB's materials for the panel included a proposal for mandatory submission of consumer financial service arbitration claims and awards that the CFPB could publish online.[12] The proposal point to several potential uses of the collected data, including continually monitoring arbitrations, identifying trends in arbitration proceedings, and assisting the CFPB and the public "in identifying potentially problematic business practices that harm consumers."[13] While claims and award documents undoubtedly contain valuable information that could be used to pursue these analyses, there is much work to be done after the production of these documents to allow for systematic analysis of the documents' content.

The arbitral archive should be easy to mobilize for both legal and economic analysis and must provide a balance between submission burden and privacy considerations across consumer financial service providers, consumers, and arbitration administrators. The intricacy and scope of this undertaking is not lost on users of other systems that aggregate documents and data across various systems,

---

[10] From a practical standpoint, the contours of the CFPB's arbitral publication system may change the size of the data collection, but are unlikely to affect its structure.

[11] *See* OUTLINE OF PROPOSALS, *supra* note 1, at 20.

[12] *See id.* (discussing the possibility of publishing claims on the CFPB's website).

[13] *Id.*

such as Public Access to Court Electronic Records (PACER),[14] either directly or through third-party search functions such as LexisNexis[15] or PacerPro.[16]

The process of preparing data for analysis is referred to in this article as "data building." The approach to a data build depends on the techniques that will be used to analyze the data. Previous studies of arbitration provide valuable insight into how arbitral claim and award data might be used in both qualitative and quantitative applications.[17]

1. Classification and Categorization

A basic necessity for any archive is a search function based on document content. Implementing a search function that is more sophisticated than a text match (or, realistically, a text match to optical character recognition text from a PDF) requires classification of content. To the extent that multiple terms may be used for the same concept or there is an overarching category not mentioned verbatim in the document, these themes might not be retrieved. For example, an arbitral claim related to payday lending may never reference "alternative financial services," "non-bank services," or "small-dollar lending." The complexity of defining categories in the CFPB's potential arbitral archive is compounded by its coverage of a variety of products and services, including checking accounts, student loans, credit reporting, credit and debit cards, payday loans, medical debt, international money transfers, and consumer deposit accounts.[18] In a particular market, this coverage may extend to multiple points that involve consumers in the value chain. For example, the CFPB's jurisdiction in credit card markets extends to credit card advertising, the underwriting and card issuance process, servicing, any related credit reporting, and debt collection.[19] Categorization of claims and awards would streamline searches for similar cases based on

---

[14] PUB. ACCESS TO CT. ELECTRONIC RECS., U.S. COURTS, https://www.pacer.gov/ [https://perma.cc/DZU6-E53P].

[15] LEXISNEXIS®, http://lexisnexis.com [https://perma.cc/QNA5-AHHG].

[16] PACERPRO, https://www.pacerpro.com/ [https://perma.cc/M5YA-D2HZ].

[17] *See infra* Part II.A.I.

[18] *See* OUTLINE OF PROPOSALS, *supra* note 1 at 22–23 (listing products and providers potentially subject to the requirements).

[19] *See* 12 U.S.C. § 5481(12) (2012) (enumerating consumer laws subject to the CFPB's jurisdiction, which include the Fair Credit Reporting Act, the Fair Debt Collection Practices Act, and the Truth in Lending Act).

defendant, product market, and a range of other factors. Such categorization could make engaging with precedent, as is common in labor arbitration, efficient.[20]

The benefit of well-defined categories is illustrated by the Financial Industry Regulatory Authority (FINRA) dispute resolution statistics, which include case filings by controversy type, security type, and open/close status.[21] These categories make it possible for FINRA to produce consistent historical statistics to examine broad arbitration and mediation trends such as the number of cases filed in a category.[22] FINRA data inherits this categorization from structure of its Online Arbitration Claim Filing System through a series of drop-down boxes and radio buttons.[23] Given the challenge of collecting and harmonizing data from multiple administrators, it is unsurprising that most empirical academic studies of arbitration rely on awards from a single administrator.[24] Since consumer financial services arbitration is handled by multiple arbitration administrators,[25] categorization will have to be harmonized across administrators or completed by a third party—likely the CFPB—after the documents are submitted.

2. Bias Analysis

The CFPB specifically emphasizes that claim and awards data might be used in diagnosing whether an arbitrator exhibits bias:

---

[20] *See* W. Mark C. Weidemaier, *Judging Lite: How Arbitrators Use and Create Precedent*, 90 N.C. L. REV. 1091, 1102 (2012) ("Legal publishers like the BNA have published labor arbitration awards—albeit selectively—for many decades, and reference texts attempt to distill the rulings of labor arbitrators into a coherent set of principles to inform future disputes.").

[21] *Dispute Resolution Statistics*, FIN. INDUSTRY REG. AUTH., http://www.finra.org/arbitration-and-mediation/dispute-resolution-statistics#arbitrationstats [http://perma.cc/9VFG-NJ3B].

[22] *See Online Claim Filing,* FIN. INDUSTRY REG. AUTH., https://www.finra.org/arbitration-and-mediation/online-claim-filing [https://perma.cc/HZT2-PHTQ].

[23] *See id.*

[24] *See, e.g.,* CONSUMER FIN. PROT. BUREAU, ARBITRATION STUDY: REPORT TO CONGRESS, PURSUANT TO DODD-FRANK WALL STREET REFORM AND CONSUMER PROTECTION ACT § 1028(A) (2015), http://files.consumerfinance.gov/f/201503_cfpb_arbitration-study-report-to-congress-2015.pdf [https://perma.cc/9ADU-VE8V] [hereinafter ARBITRATION STUDY].

[25] *See id.* at 35 (listing the American Arbitration Association, JAMS, and the National Arbitration Forum as arbitrators of consumer financial disputes).

> The [CFPB] believes that there is a potential for consumer harm if arbitration agreements were to be administered by biased administrators (as was alleged in the case of the NAF) or individual arbitrations were otherwise conducted in an unfair manner. Thus the [CFPB] is considering a limited intervention that would serve to deter the emergence of such unfair arbitrations and also to shed sunlight on any unfairness that might emerge, while at the same time would impose minimal regulatory burdens on current arbitration activity.[26]

A significant issue with this approach is that win-loss data bias measures are difficult to interpret, particularly once they are used as a diagnostic. This interpretation problem is illustrated by three types of measures: (1) measures that use small samples, (2) measures that omit variables, and (3) measures that reflect adverse incentives created by the use of the data itself.

First, the power of statistical analysis in diagnosing bias depends on the number of observations available for analysis. The larger the number of observations is for any given analysis, the higher the power of the test is.[27] The CFPB's arbitration study dataset contains two affirmative Electronic Fund Transfer Act (EFTA) claims filed in 2010–2011 that were resolved by arbitrators, both of which resulted in affirmative consumer awards.[28] Suppose that each of the two arbitrations were conducted by a different arbitrator, and that one resulted in a consumer award, while the other did not. The arbitrator's records would be 100 percent consumer awards and 0 percent consumer awards, respectively. However, standard statistical tests do

---

[26] OUTLINE OF PROPOSALS, *supra* note 1, at 19.

[27] *See* RICHARD J. LARSEN & MORRIS L. MARX, AN INTRODUCTION TO MATHEMATICAL STATISTICS AND ITS APPLICATIONS 384 (Kathleen Boothby Sestak et. al. eds., 3rd ed. 2001) (defining statistical power as the probability that we accept the null hypothesis of no statistical difference between groups when there actually is a difference between groups: "it represents the ability of the decision rule to 'recognize' (correctly) that H_0 is false."); *id.* at 388–89 ("[T]he sample size is the parameter that researchers almost invariably turn to as the mechanism for ensuring that a hypothesis test will have a sufficiently high power against a given alternative.").

[28] ARBITRATION STUDY, *supra* note 24, § 5 at 49 fig.6 (charting the "substantive outcome by claim type for affirmative claims in arbitrator resolved disputes, disputes filed in 2010–2011") (emphasis omitted).

not reject the equality of outcomes even when only one observation for each arbitrator is observed.[29] These misleading outcomes are particularly important when a data search is restricted to particular types of arbitrations or arbitrations that occur within a particular interval.[30]

Even when more arbitrations are observed for each arbitrator for each type of arbitration, the data collected about each claim can affect the analysis. If arbitrators are not assigned randomly to cases, then information observed in the data may not fully capture the differences between arbitrations, even with sophisticated techniques like regression analysis. Parties must mutually agree on an arbitrator, which they may choose based on whether the arbitrator's qualifications match the needs of the case.[31] Consider two arbitrators that have identical characteristics in the data but who arbitrate cases that differ in ways not captured in the data. Suppose Arbitrator A specializes in cases that should result in consumer awards 40 percent of the time based on merits and Arbitrator B specializes in cases that should result in awards 60 percent of the time based on merits. This specialization may be based on a variety of factors, such as previous experience with the product or legal concepts related to the case. Accordingly, even though both arbitrators are neutral, Arbitrator B appears more favorable to consumers than Arbitrator A due to arbitrator assignment.

---

[29] *See* LARSEN & MARX, *supra* note 25, at 506 (applying the formula to test the equality of the proportion of successes for two Bernoulli trials). We compute a z-score (indicating how many standard deviations an element is from the mean) of 1.41, which fails to reject the hypothesis that the proportions of arbitrations that result in a consumer award is different between the two arbitrators at a 5 percent level of significance.

[30] *See* ARBITRATION STUDY, *supra* note 24, § 9 at 11 (stating that the CFPB's analysis was limited because it was not able to identify all of the public and private actions and was not able to search all jurisdictions).

[31] *Arbitration: Arbitration Process—Arbitrator Selection*, AM. ARBITRATION ASS'N, https://www.adr.org/aaa/faces/services/disputeresolutionservices/arbitration?_afrWindowId=19lgas6gl_50&_afrLoop=1339953530625928&_afrWindowMode=0&_adf.ctrl-state=19lgas6gl_53 [https://perma.cc/SAR6-EPRZ] ("The Parties' criteria are used to identify neutrals with qualifications that match the needs of the Case. Once the parties agree on the neutral, the arbitration proceedings may begin.").

A diagnostic for neutrality may be benchmarked against an absolute proportion of consumer awards or other arbitrators.[32] However, if arbitrators think that their selection depends on their appearance of neutrality and understand how neutrality is measured, then this information can potentially create adverse incentives even with detailed data collection.[33] Arbitrator A in the example above might take steps to manage her reputation for neutrality.[34] For example, despite a lack of familiarity with the subject matter, she might try to get selected for more cases that should objectively result in consumer awards. If the neutrality measure is relative to Arbitrator B, the two might be able to "trade" cases in order to converge on similar award records. Furthermore, Arbitrator A might modify award decisions in order to balance her record. As Klement and Neeman point out, "[s]ince the only way an arbitrator can establish a reputation for being impartial is by avoiding a series of decisions that might seem biased against a specific group, she might want to make an incorrect decision when a correct decision may raise the suspicion that she is biased."[35]

Given the CFPB's assertion that arbitration data might be used to identify arbitrator biases, it should address issues arising from small sample sizes, unconsidered variables, and arbitrator incentives in designing and implementing an archive.

B. Privacy Considerations Specific to Consumer Financial Arbitral Claims and Awards

The CFPB's SBREFA report states, "[b]efore collecting or publishing any arbitral claims or awards, the [CFPB] would ensure that these activities comply with privacy considerations."[36] The report

---

[32] *See generally* William Park, *Arbitrator Bias* 63 (Bos. Univ. Sch. of Law, Working Paper No. 15-39, 2015) (asserting that arbitrators can be assessed based on their relative "relationships and predispositions").

[33] *See generally* Alon Klement & Zvika Neeman, *Does Information About Arbitrators' Win/Loss Ratios Improve Their Accuracy?*, 42 J. LEGAL STUD. 369, 373 (2013) (arguing that providing litigants information on arbitrators' past decisions may lead to negative incentives for arbitrators as it may cause arbitrators to decide in a way that helps the arbitrator avoid appearing partial to certain types of parties).

[34] *Id.*

[35] Klement & Neeman, *supra* note 33, at 373.

[36] OUTLINE OF PROPOSALS, *supra* note 1, at 20.

does not, however, identify the privacy considerations at issue.[37] The CFPB must first consider how to define the arbitration archival privacy standards, and then clearly articulate any proposed privacy requirements in order for the public and covered entities to evaluate the arbitral publication system.

In the United States, privacy protections generally depend on context; letters from nursing home residents are generally protected,[38] while letters from deployed soldiers historically have been censored.[39] No law expressly articulates unique privacy considerations applicable to the collection and dissemination of arbitral claims and awards.[40] Therefore, the CFPB has discretion to articulate relevant privacy considerations,[41] and those considerations will affect the costs and benefits of the arbitral publication system.[42] For example, consumers may be particularly protective of information about their consumer financial services arbitration experiences because of the stigma

---

[37] *See generally id.*

[38] *See, e.g.*, ARIZ. ADMIN. CODE § R9-10-711 (2014), https://apps.azsos.gov/public_services/Title_09/9-10.pdf [http://perma.cc/DZ3U-UVH8] (asserting that residents are "[not] to be prevented or impeded from exercising the resident's civil rights unless the resident has been adjudicated incompetent or a court of competent jurisdiction has found that the resident is unable to exercise a specific right or category of rights").

[39] *See, e.g.*, Trading with the Enemy Act of 1917, Pub. L. No. 65-91, § 3(d), 40 Stat. 411, 413 (1917) (stating President may censor communication between United States and any foreign country for the public safety).

[40] *See generally* Lisa Bench Nieuwveld, *CONFIDENTIALITY: Not to Be Overlooked When Drafting the Arbitration Clause*, KLUWER ARB. BLOG (May 17, 2012), http://kluwerarbitrationblog.com/2012/05/17/confidentiality-not-to-be-overlooked-when-drafting-the-arbitration-clause/ [https://perma.cc/EET2-9DBJ] ("The US Federal Arbitration Act does not address confidentiality, although courts generally recognize this as important to arbitration."). Laws generally applicable to the collection and dissemination of data, such as the Privacy Act of 1974 and Section 1022(c)(8) of the Dodd-Frank Act, articulate privacy considerations. We assume that the CFPB's statement regarding compliance with privacy considerations refers to considerations unique to an arbitral publication system, beyond the general considerations applicable to government collection and dissemination of data.

[41] *See also* Michelle Frasher, *Adequacy Versus Equivalency: Financial Data Protection and the U.S.-EU Divide*, 56 BUS. HORIZONS 787, 793 (2013) (discussing the CFPB's jurisdiction over privacy oversight).

[42] *See infra* notes 88–109 and accompanying text.

associated with bankruptcy or debt.[43] Accordingly, it is important to understand what privacy considerations the CFPB is contemplating for its proposed policy.

Existing privacy regulations specify varying standards for protected private information. Agencies commonly rely on the Government Accountability Office (GAO) definition of personally identifiable information (PII),[44] which is defined as "any other information that is *linked or linkable* to an individual, such as medical, educational, financial, and employment information." (emphasis added).[45] These principles are familiar; the CFPB already considers them with respect to its own operations, such as the consumer complaint database[46] and Freedom of Information Act requests.[47] But these principles, which are intended to appropriately maximize

---

[43] *See generally* John Gathergood, *Debt and Depression: Causal Links and Social Norm Effects*, 122 ECON. J. 1094, 1109, 1094–1114 (2012) (discussing "the impact of social stigma arising from problem debt" on the social norm effect "present in the relationship between problem debt and psychological health"); Scott Fay et al., *The Household Bankruptcy Decision,* 92 AM. ECON. REV. 706–18 (2002) (discussing the bankruptcy stigma and its economic implications).

[44] *See, e.g.*, Memorandum from Clay Johnson III, Deputy Dir. for Mgmt., Office of Mgmt. & Budget, Exec. Office of the President to Heads of Exec. Dep't & Agencies (May 22, 2007), https://www.whitehouse.gov/sites/default/files/omb/memoranda/fy2007/m07-16.pdf [https://perma.cc/AF2A-H8DF] (finding that agencies are governed by the Privacy Act, in which "personally identifiable information" is defined with a similar definition to the GAO's definition).

[45] U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-08-536, PRIVACY: ALTERNATIVES EXIST FOR ENHANCING PROTECTION OF PERSONALLY IDENTIFIABLE INFORMATION 1, n.1 (2008), http://www.gao.gov/new.items/d08536.pdf [https://perma.cc/PZ3F-SBAG].

[46] *Consumer Complaint Database*, CONSUMER FIN. PROT. BUREAU, http://www.consumerfinance.gov/data-research/consumer-complaints/ [https://perma.cc/H6T8-BCTS] (stating "we publish the consumer's description of what happened if the consumer opts to share it and after taking steps to remove personal information" on a page for the CFPB Consumer Complaint database).

[47] *FOIA Requests*, CONSUMER FIN. PROT. BUREAU, http://www.consumerfinance.gov/foia-requests/ [https://perma.cc/HYJ6-95M5] (mentioning the Privacy Act, which governs the use of personally identifiable information, on the CFPB's Freedom of Information Act page).

government transparency, are necessarily broad.[48] A privacy rule that follows the same approach would maximize redaction and limit the utility of an arbitral publication system.

While the GAO's broad definition of PII might not satisfy the goals of an arbitral publication system, the financial service industry defines PII more narrowly. Regulation P, promulgated under Section 504 of the Gramm-Leach-Bliley Act (GLB), defines personally identifiable financial information to include any information "about a consumer resulting from any transaction involving a financial product or service" between a covered person and a consumer, or information that a covered person "obtains about a consumer in connection with providing a financial product or service to that customer."[49] Without the GAO's "linked or linkable" language, this definition is narrower in scope and would be less burdensome to satisfy. Nevertheless, the GLB language is sufficiently protective that it has the potential to hamper the CFPB's transparency goals.

State law presents another set of privacy issues the CFPB should consider in implementing an arbitration archive. Several states have financial privacy protections similar to, and in some cases stronger than, those provided by GLB.[50] The CFPB must decide whether to incorporate state law requirements into its arbitrational archive privacy protection requirements. Further, it must decide whether the entity responsible for redacting private information from the archives should tailor these redactions to the consumer's state of residence.

The CFPB may also want to consider state requirements related to the confidentiality of the arbitration proceedings themselves. While several states limit the admissibility of arbitration information in legal proceedings, Missouri completely prohibits the disclosure of

---

[48] *See Transparency of Federal Data*, U.S. GOV'T ACCOUNTABILITY OFFICE, http://www.gao.gov/key_issues/transparency_federal_data/issue_summary [https://perma.cc/N5JM-JZQX] ("Public access to reliable and complete federal financial and performance data can foster transparency, improve oversight, and enhance public participation.").

[49] Privacy of Consumer Financial Information (Regulation P), 12 C.F.R. § 1016.3(q)(1) (2016); Gramm-Leach-Bliley Act, 15 U.S.C. § 6804(a)(1)(A) (2012).

[50] *See*, *e.g.*, CAL. FIN. CODE §§ 4050–60 (West 2015) (stating that the protections of the GLB are inadequate for California residents and so the state imposes stricter privacy standards); CONN. GEN. STAT. ANN. §§ 36a-41–45 (West 2011) (describing privacy standards that are similar to the GLB).

information related to arbitration.[51] Texas and Arkansas similarly prohibit arbitration-related disclosures, and require an in-camera judicial proceeding to determine whether protected information can be disclosed in the event the confidentiality requirements conflict with other legal requirements.[52] As in other contexts, the CFPB should consider whether to tailor arbitral disclosure requirements in light of the relevant goals and considerations applicable in different states.[53]

In some cases, the CFPB should also consider privacy implications beyond those generally applicable to financial organizations. Special privacy considerations are imposed under the Health Insurance Portability and Accountability Act (HIPAA).[54] Under HIPPA, protected health information includes any information that relates to "the provision of health care to an individual" and "the past, present, or future payment for the provision of health care to an individual."[55] This type of information may be included in arbitral claims and awards. Medical point-of-sale lending, such as loans for orthodontia, cosmetic surgery, or laser eye surgery, is an increasingly

---

[51] MO. ANN. STAT. § 435.014 (2010) (stating that arbitrators, conciliators, mediators or their agents may not disclose "any matter disclosed in the process of setting up or conducting the arbitration" and that "any communication relating to the subject matter of such disputes" is not obtainable in discovery).

[52] TEX. CIV. PRAC. & REM. CODE ANN. § 154.073 (requiring all communications relating to a civil or criminal dispute made in an alternative dispute resolution procedure to be confidential, or, if there is a conflict with another legal requirement, that the "issue of confidentiality" be "presented to the court . . . in camera"); ARK. CODE ANN. § 16-7-206 (mandating the confidentiality of communications in dispute resolution procedures, or, if confidential information conflicts with other legal requirements, the issue must be presented to a court "in camera" to determine if the material is subject to disclosure).

[53] U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-14-758, CONSUMER FINANCIAL PROTECTION BUREAU: SOME PRIVACY AND SECURITY PROTECTION PROCEDURES SHOULD CONTINUE BEING ENHANCED 1 (2014), http://www.gao.gov/assets/670/666000.pdf [https://perma.cc/XVB9-9XJB]. (describing how the CFPB has adopted procedures for privacy and information security, including adhering to the Federal Information Security Management Act's standards and following guidelines promulgated by the National Institute of Standards and Technology).

[54] *See generally* Health Insurance Portability and Accountability Act of 1996, 42 U.S.C. § 1320d (2012).

[55] §1320d(4)(A).

important part of the consumer financial services market.[56] While medical financing is sometimes provided by a third-party financial services company, in many cases, the medical provider directly provides patient financing.[57] Similar issues exist in medical debt collection, which is another area of concern to the CFPB.[58] The medical financing industry should carefully consider the CFPB's proposal and be prepared to work with the CFPB on issues arising in this context.[59] Determining what information should be redacted will likely require a qualitative analysis of the information included in the arbitral claim and award. Even if the names of consumers are removed, the CFPB must implement a procedure for determining whether there is a reasonable basis to believe the information included in an arbitral claim or award can be used to identify the individual.

The CFPB must consider practical healthcare privacy issues beyond the determination of what information to redact. U.S. Department of Health and Human Services (HHS) regulations impose security, notification, and privacy requirements on persons in possession of protected health information.[60] Even if the CFPB is not subject to these rules, the CFPB must determine if it is willing to provide a similar level of protection for sensitive health information. Additionally, if the CFPB imposes a data collection requirement and requires exams for compliance with the requirement, it must determine procedures for what the CFPB will do if it discovers violations of

---

[56] In 2014, GAO estimated that more than four million consumers use medical financing. U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-14-570, CONSUMER FINANCE: CREDIT CARDS DESIGNED FOR MEDICAL SERVICES NOT COVERED BY INSURANCE 1 (2014), http://www.gao.gov/assets/670/664244.pdf [https://perma.cc/XEU4-E5RD] (using the term "medical credit cards" to "refer collectively to financial products—including revolving credit lines and installment loans—that are designed specifically to finance health care services not covered by health insurance").

[57] *See id*. ("Financial institutions offer medical credit cards through participating providers to consumers (patients).").

[58] *See generally* KENNETH P. BREVOORT & MICHELLE KAMBARA, CONSUMER FIN. PROT. BUREAU, DATA POINT: "MEDICAL DEBT AND CREDIT SCORES (2014), http://files.consumerfinance.gov/f/201405_cfpb_report_data-point_medical-debt-credit-scores.pdf [https://perma.cc/5RG7-NSPB].

[59] § 1320d-5 (discussing protections available to consumers regarding medical debt collections).

[60] 45 C.F.R. §§ 164.302–318 (2012) (describing security requirements); §§ 164.400–414 (describing notification requirements); §§ 164.500–534 (describing privacy requirements).

HHS's security, notification, or privacy regulations during an exam. The CFPB should consider whether it is prepared to work with HHS or state Attorneys General to address such violations.[61]

Special privacy considerations also apply to students. While most education financing is provided through private student lenders or the Department of Education, nonprofit public and private universities lend directly to students.[62] The Family Educational Rights and Privacy Act (FERPA) imposes privacy requirements on such lending institutions.[63] Rules promulgated under FERPA protect PII and define it expansively to include "information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person in the school community, who does not have personal knowledge of the relevant circumstances, to identify the student with reasonable certainty."[64] If a dispute over a loan from an educational institution to a student leads to arbitration, information included in educational records may be included in arbitral claims and awards. Lawsuits between nonprofit educational institutions and indebted students are increasingly common, and arbitration may become the preferred forum over time.[65] Redacting some types of FERPA PII, such as the student's name or personal identifier, might be straightforward, but redacting linked or linkable information would involve a qualitative analysis of the facts and circumstances. However, the prospect of imposing new compliance costs on these educational

---

[61] 42 U.S.C. § 1320d-5 (2012) (providing HHS and state Attorneys General with enforcement authority).

[62] *See, e.g., Institutional Loans*, LOYOLA MARYMOUNT UNIV., http://financialaid.lmu.edu/prospective/faq/institutionalloans/ [https://perma.cc/BQA9-KPDS] (exemplifying a nonprofit private post-secondary institution that offers loans directly to students); *Institutional Loan Program*, UNIV. OF CAL., BERKELEY, http://studentbilling.berkeley.edu/InstitutionalLoanProgram.htm [https://perma.cc/B4PH-28QC] (providing an example of direct financing at a public post-secondary institution).

[63] Family Educational Rights and Privacy Act of 1974, 20 U.S.C. § 1232g (2012) (governing access and review of educational records).

[64] 34 C.F.R. § 99.3(f) (2016).

[65] *See* Janel Lorin, *Yale Suing Former Students Shows Crisis in Loans to Poor*, BLOOMBERG (Feb. 5, 2013, 12:01 AM), http://www.bloomberg.com/news/articles/2013-02-05/yale-suing-former-students-shows-crisis-in-loans-to-poor [https://perma.cc/DGF5-V8F6] (describing the trend of suits brought by universities against former students for unpaid loans not made by the federal government but the universities themselves).

institutions, which may be passed on to indebted students, warrants serious consideration.[66]

These complexities illustrate the CFPB's core problem—protecting privacy requires substantial efforts. If the CFPB takes an under-inclusive approach and requires redaction of only a few items, costs on industry will be minimized and the data disclosed to the public will be maximized, but the risk of consumer harm will increase. If the CFPB takes an over-inclusive approach the risk of consumer harm will decrease, but industry costs will increase and the published data will not maximize the benefits of public access.[67]

Given these challenges, the CFPB should consider whether the proposed privacy considerations need to be defined based on type of institution involved, or type of consumer affected, or nature of the dispute.[68] No matter the approach proposed, the clearer the CFPB articulates the privacy standard, the easier it will be for the public to evaluate the utility the arbitral publication system.

III. Implementation Considerations

A. Ex-Post Hand Coding: The Simplest Approach?

Standardization and categorization of data is fundamental to any analysis. The CFPB chose a simple hand-coding approach in its Arbitration Study.[69] The CFPB's SBREFA materials acknowledge that "[t]he [CFPB] believes that the Study provides the most comprehensive data on individual consumer financial arbitration frequency and outcomes to date."[70] CFPB staff should be lauded for this labor-intensive undertaking. They manually coded all non-class consumer awards from American Arbitration Association (AAA) case management records that were received from January 2010 to

---

[66] Although the pass through of costs may not be prevalent in other markets, the potential inelasticity of demand for student debt may present a special case.

[67] *See* OUTLINE OF PROPOSALS, *supra* note 1, at 19–20 (explaining that privacy risk to consumers should be balanced with the need to meet information objectives).

[68] *See id.* at 5 (describing privacy considerations can vary depending on the purpose of the inquiry).

[69] ARBITRATION STUDY, *supra* note 24 (describing simple counting as the technique used in a study of arbitration clause administrators)

[70] OUTLINE OF PROPOSALS, *supra* note 1, at 19.

February 2013.[71] The data review procedures are documented in Appendix B of the report,[72] and include codes constructed at the researchers' discretion.[73] This type of coding is sometimes used in coding qualitative data,[74] and often requires that the coder be knowledgeable about the material being coded.

Academic studies of arbitration have taken data-building approaches similar to the CFPB Arbitration Study.[75] While the CFPB study involved hand-coding 1,241 consumer arbitrations related to checking accounts, credit cards, and payday loans,[76] larger scale arbitration data building exercises have been undertaken to study both securities and labor arbitration.[77] In its proposal, the CFPB draws an analogy to FINRA publication of awards and AAA publication of employment awards,[78] and states that arbitral claim and award data "would also be helpful to the Bureau, consumers, companies, and possibly to other regulatory entities and academics who study

---

[71] ARBITRATION STUDY, *supra* note 24, at 136 (Appendix B).

[72] *Id.* (describing data review procedures used by the CFPB).

[73] *Id*. at 140 (Appendix B) (explaining that the study made assumptions to determine which party filed for arbitration, and then included that information in the analysis).

[74] *See generally* LEE EPSTEIN & ANDREW D. MARTIN, AN INTRODUCTION TO EMPIRICAL LEGAL RESEARCH (2014) (discussing use of data in legal research, including a discussion of coding in Chapter 5).

[75] *See generally* ARBITRATION STUDY, *supra* note 24, at 134 (Appendix B) (explaining the data process used, which was hand coding, and the results of the study).

[76] *Id*. at 136 (Appendix B) (explaining the process for obtaining these claims from AAA and sorting them by product market).

[77] Seth E. Lipner, *Expungement of Customer Complaint CRD Information Following Settlement of a FINRA Arbitration*, 19 FORDHAM J. CORP. & FIN. L. 57, 91 (2013) (explaining that Lipner performed a text search for the term "expungement" in FINRA records for the first six months of 2013 to understand the effects of expungement of consumer complaints following the settlement of a FINRA arbitration claim); Alexandre Mas, *Pay, Reference Points, and Police Performance* 783–821 (Quarterly Journal of Econ., Working Paper No. 12202, 2006); Orley Ashenfelter & Gordon B. Dahl, *Bargaining and the Role of Expert Agents: An Empirical Study of Final-Offer Arbitration*, 94 REV. ECON. & STAT. 116–32 (2010) (discussing hand-coded 1978–1996 New Jersey municipality and police bargaining unit arbitration).

[78] OUTLINE OF PROPOSALS, *supra* note 1, at 20–21 (explaining that these are the "two main administrators" and do not "require publication of claims or awards in matters between consumers and providers of financial services").

consumer finance."[79] While the FINRA and AAA awards mentioned above are published, researchers hand-coded award data before performing qualitative analysis.[80] In 2007, the Cornell Industrial and Labor Relations School purchased and hand-coded 3,200 arbitration awards issued by FINRA and its predecessors, which various authors then used in multiple academic paper.[81] Subsequent papers have made use of another hand-coding of FINRA records from November 1992 to December 2016.[82] Similar hand-coding projects were undertaken in research related to employment disputes in securities[83] and labor and employment arbitration.[84]

Even when data is published as a spreadsheet by the arbitration administrator, as AAA did for its consumer arbitration

---

[79] *Id.* at 21 (explaining how these reports and information would be beneficial to multiple parties in arbitration situations).

[80] *See generally* J. Ryan Lamare & David B. Lipsky, *Employment Arbitration in the Securities Industry: Lessons Drawn from Recent Empirical Research*, 35 BERKELEY J. EMP. & LAB. L. 113, 119 (2014) (explaining the data was "cleaned and coded" before it was analyzed).

[81] *Id.* at 119 (analyzing awards of various types, including: discrimination, breach of contract, compensation, defamation, and wrongful termination). *See* David B. Lipsky et al., *The Arbitration of Employment Disputes in the Securities Industry: A Study of FINRA Awards, 1986–2008*, 65 DISP. RESOL. J. 12, 54–57 (2010) (explaining the process used by the Cornell Industrial and Labor Relations School to gather and analyze the awards data).

[82] *See, e.g.*, Stephen J. Choi et al., *The Influence of Arbitrator Background and Representation on Arbitration Outcomes*, 9 VA. L. & BUS. REV. 43, 63–82 (2014) (conducting a series of empirical tests using FINRA records); Stephen J. Choi & Theodore Eisenberg, *Punitive Damages in Securities Arbitration: An Empirical Study*, 39 J. LEGAL STUD. 497 (2009) (using a data set of over 6,800 arbitration awards from FINRA records to empirically study punitive damage awards).

[83] *See generally* Lipner, *supra* note 77, at 91–95 (performing a text search for the term "expungement" in FINRA records for the first six months of 2013 to understand the effects of expungement of consumer complaints following the settlement of a FINRA arbitration claim).

[84] *See generally* Alexandre Mas, *Pay, Reference Points, and Police Performance*, 121 Q.J. ECON. 783, 788–93 (2006) (examining data on final offer arbitration for police unions); Orley Ashenfelter & Gordon B. Dahl, *Bargaining and the Role of Expert Agents: An Empirical Study of Final—Offer Arbitration*, 94 REV. ECON. & STAT. 116, 117 (2012) (outlining hand-coded 1978–1996 New Jersey municipality and police bargaining unit arbitration).

statistics pursuant to certain state statutes,[85] the coding may not reflect the information necessary for the agency's purposes. For example, consumer financial services arbitration can be isolated, but the product type, such as brokerage product versus educational lending, is not coded.[86] Getting product-type information for these records would require additional coding involving researcher discretion. Specific research projects require less planning and dataset demands than ongoing government data collection and dissemination.[87] Even a well-resourced retrospective hand-coding approach may not achieve the CFPB's intended goals if implemented in lieu of a well-defined, forward-looking standard for data reporting. Accordingly, a standardized ex ante approach to data collection may be best suited for the CFPB's goals.

B. Costs and Benefits of Data Collection and Maintenance

The aggregation, standardization, and dissemination of public data is a major undertaking that involves both public and private sector participants.[88] Beyond the participation of firms required to submit the

---

[85] *Government & Consumer: Consumer Arbitration Statistics*, AM. ARBITRATION ASS'N, https://www.adr.org/aaa/faces/aoe/gc/consumer/consumerarbstat?_afrWindowId=14fxmwf2i0_201&_afrLoop=1027970136106473&_afrWindowMode=0&_adf.ctrl-state=14fxmwf2i0_204 [https://perma.cc/3TF6-TDF9]; *see also* Alexander J. S. Colvin & Mark D. Gough, *Individual Employment Rights Arbitration in the United States*, 68 INDUS. & LAB. REL. REV. 1019, 1026–27 (2015) (using data collected in spreadsheet format by AAA).

[86] *See, e.g.*, AM. ARBITRATION ASS'N, PROVIDER ORGANIZATION REPORT, https://www.adr.org/cs/idcplg?IdcService=GET_FILE&dDocName=ADRSTAGE2041881&RevisionSelectionMethod=LatestReleased [https://perma.cc/SC79-7RTQ] (coding consumer arbitration statistics without coding product type).

[87] *See generally* OFFICE OF FIN. RESEARCH, VIEWPOINT: DEVELOPING BEST PRACTICES FOR REGULATORY DATA COLLECTIONS (2016) https://www.financialresearch.gov/viewpoint-papers/files/OFRvp-2016-01_Best-Practices-Data-Collection.pdf [https://perma.cc/W3Q2-9WVZ] (detailing best practices for ongoing data collection and dissemination).

[88] *See generally* Lawrence H. Summers, *Data Collection is the Ultimate Public Good*, WASH. POST (Apr. 4, 2016), https://www.washingtonpost.com/news/wonk/wp/2016/04/04/larry-summers-data-collection-is-the-ultimate-public-good/ [https://perma.cc/UHN7-X2RC].

data, there are also industries created around reformatting, retrieving, and validating data from government sources.[89] Both government and private sector data disseminators must assure that their product is quality controlled and consistent over time.[90] These quality control measures include thorough documentation, particularly of any efforts to harmonize the data across multiple sources.[91] In the case of the CFPB's proposal to collect arbitral claims and awards, this requires, at a minimum, combining data from AAA and JAMS, Inc. (JAMS).[92]

1. Potential Benefits

In order to benefit from the use of an arbitral archive, the CFPB must decide who will analyze its consumer arbitration data, and how. Various examples of successful ongoing government data projects build on the submission of administrative data. For example, the Bureau of Labor Statistics Quarterly Census of Employment and Wages makes it possible to compute quarterly employment statistics at a local level based on information submitted by employers.[93] The U.S. Energy Information Administration incorporates state agency data through third-party sources, and the Department of Interior uses that

---

[89] *See generally* David Robinson et al., *Government Data and the Invisible Hand*, 11 YALE J.L. & TECH. 160, 161 (2009) ("Private actors, either nonprofit or commercial, are better suited to deliver government information to citizens and can constantly create and reshape the tools individuals use to find and leverage public data.").

[90] *See id.* at 165 ("[T]he desire to increase data quality by adopting a uniform method of identifying the recipients of federal funds has led to proposed amendments to the original legislation, aimed at improving data accuracy and standardization across agencies.").

[91] *See generally* STAT. CAN., STATISTICS CANADA'S QUALITY ASSURANCE FRAMEWORK 12-586-XIE (2002), http://www.statcan.gc.ca/pub/12-586-x/12-586-x2002001-eng.pdf [https://perma.cc/5GX7-8ENM] (outlining official Statistics Canada's framework for data quality includes best practices and examples of documentation and quality control).

[92] *See* ARBITRATION STUDY, *supra* note 24, § 2, at 35 (observing that JAMS and AAA are sole arbitration opinion in some arbitration clauses).

[93] *See Quarterly Census of Employment and Wages*, BUREAU OF LAB. STATISTICS OF THE U.S. DEP'T OF LAB., http://www.bls.gov/cew/home.htm [https://perma.cc/5A3M-4JUG] ("[P]ublish[ing] a quarterly count of employment and wages reported to employers covering 98 percent of U.S. jobs, available at the county, MSA, state and national levels by industry.").

data to project monthly crude oil production estimates.[94] The Department of Education's National Center for Educational Statistics has been fielding the National Postsecondary Student Aid Survey that combines administrative data on postsecondary transcripts, financial aid, and test scores in addition to a survey component.[95] These are examples of successful government efforts to overcome coordination problems in assembling data from disparate sources. These data collections and their attendant reports, and publicly available data comprise a fountain of information that is used in a variety of business, policy, and research applications.[96]

Producing high-quality data for public consumption is a major undertaking. After the data collection is designed, dedicated staff and information technology resources are critical to ensuring that accuracy and privacy standards are met and reports are produced.[97] Additionally, if arbitration data is made publically available, the CFPB must be able to answer user questions about technical matters, such as the definition of fields and any methods used to suppress confidential information, should also be provided.

To the extent that a government data product may not be readily available in the format that the end user requires for analysis,

---

[94] *See generally* U.S. ENERGY INFO. ADMIN., METHODOLOGY FOR MONTHLY CRUDE OIL ESTIMATES (2015), http://www.eia.gov/petroleum/supply/monthly/pdf/crudemeth.pdf [https://perma.cc/C32Z-YA4E] (estimating monthly crude oil production and explaining the process by which estimates are calculated, including the data sources, estimation techniques, and role of expert judgment).

[95] *See National Postsecondary Student Aid Study—About NPSAS*, U.S. DEP'T OF EDUC., http://nces.ed.gov/surveys/npsas/about.asp [https://perma.cc/4K3J-99WN] (providing a compilation of a "comprehensive research dataset, based on student-level records, on financial aid provided by the federal government, the states, postsecondary institutions, employers, and private agencies, along with student demographic and enrollment data").

[96] *Data Impact*, OFF. CITIZEN SERV. & INNOVATIVE TECH., U.S. GEN. SERV. ADMIN., https://www.data.gov/impact/ [https://perma.cc/9HPN-47BG] ("Open government data is important because the more accessible, discoverable, and usable data is, the more impact it can have. These impacts include, but are not limited to: cost savings, efficiency, fuel for business, improved civic services, informed policy, performance planning, research and scientific discoveries, transparency and accountability, and increased public participation in the democratic dialogue.").

[97] *See generally* The Family Educational Rights and Privacy Act of 1974, 20 U.S.C. § 1232g (2012).

there are a series of third-party intermediaries that provide additional search and processing for a fee.[98] Title companies, which perform and guarantee searches of government records of real property ownership, make up one such industry.[99] For data that can be searched on a document-by-document basis, there is also a growing industry of firms that code the data so that it can be analyzed quantitatively.[100] These include EDGAR Online, which converts publicly available SEC Electronic Data Gathering, Analysis, and Retrieval (EDGAR) filings to quantitative data,[101] and FNC National Collateral Database, which collects and harmonizes information from real estate appraisal and local property assessments.[102] The existence of these services indicates that the market places value on the intermediate processing of government data. Accordingly, the CFPB must coordinate who will

---

[98] *See generally* JOHN HALTIWANGER, NAT'L. SCI. FOUND.,,,, MAKING "DRILL DOWN" ANALYSIS OF THE ECONOMY A REALITY (2010), https://www.nsf.gov/sbe/sbe_2020/2020_pdfs/Haltiwanger_John_230.pdf (highlighting applications and methodological considerations when using government and private sector data, submitted as part of the National Science Foundation SBE 2020: Future Research in the Social, Behavioral and Economic Sciences program).

[99] *See generally What Does a Title Company Do?*, ZILLOW, http://www.zillow.com/mortgage-learning/title-company/ [https://perma.cc/ND76-SXTW] ("The title company makes sure a property title is legitimate, so that the buyer may be confident that once he buys a property, he is the rightful owner of the property. To ensure that the title is valid, the title company will do a title search, which is a thorough examination of property records to make sure that the person or company claiming to own the property does, in fact, legally own the property and that no one else could claim full or partial ownership of the property.").

[100] *See, e.g.*, *Data Content Solutions*, EDGAR ONLINE, http://www.edgar-online.com/DataContentSolutions.aspx [https://perma.cc/ERD5-5WJT] (providing a database of financial and regulatory filings, and offering subscriptions enabling users to directly source the data).

[101] *Id.* (providing subscription-based access to all EDGAR filings for sourcing purposes).

[102] *See National Collateral Database*, FNC, http://www.fncinc.com/Products/ncd.aspx?ref=63 [https://perma.cc/MQB9-CU79] ("The FNC National Collateral Database has been built from data aggregated from many FNC clients—the nation's major mortgage lenders—who agreed to share their non-confidential appraisal data, so they will have access to the database. That appraisal data is blended with public record data from tax assessors and county recorders to create the most complete and timely data on residential properties available.").

process arbitration data and how it will be made available in order to reap the benefits of the arbitral archive.

2. Potential Costs

The CFPB must also consider the costs of collecting, organizing, and retaining data. Government agencies have taken different approaches to fund data collection and dissemination. One common strategy is to include data collection and processing as part of the organization's budget. For example, the Department of Energy's FY 2016 budget includes a $122 million line item for Energy Information Administration funding.[103] Alternatively, agencies including the Federal Financial Institutions Examination Council split costs between its member organizations (Federal Reserve System, Federal Deposit Insurance Corporation, Office of the Comptroller of the Currency, the CFPB, Housing and Urban Development, and mortgage insurance companies) for the production and distribution of data and reports.[104] The expenditures for the data and reports totaled $4.6 million in 2014 and $4.2 million in 2013.[105] Finally, some government entities charge the user directly for use of the system.[106] For example, PACER is available on a fee-for-service basis at a cost of $0.10 per page and $2.40 per audio file, with a $15.00 charge exemption per quarter for the indigent and certain pro-bono work.[107]

Data costs, of course, depend on what data is collected and for how long it is retained. For example, California Code of Civil Procedure § 1281.96 requires that all information related to consumer arbitration commenced on or after January 1, 2003 be retained for five

---

[103] *Budget and Performance*, U.S. ENERGY INFO. ADMIN., https://www.eia.gov/about/budget_performance.cfm [https://perma.cc/MJ4P-DJ4G] ("The fiscal year (FY) 2016 budget provides $122 million for EIA, a $5 million increase over EIA's funding in FY 2015.").

[104] *See generally* FED. FIN. INST. EXAMINATION COUNCIL, ANNUAL REPORT 2014 (2015), http://www.ffiec.gov/PDF/annrpt14.pdf [https://perma.cc/5JPB-XBHY] (reporting on the "actions the Council has taken during the year to foster communication, cooperation, and coordination to promote uniformity in the supervision of financial institutions").

[105] *Id*. at 48 (reporting on the data processing costs for 2013 and 2014).

[106] *See generally* PUB. ACCESS TO COURT ELEC. RECORDS (PACER), ELECTRONIC PUBLIC ACCESS FEE SCHEDULE (Dec. 2013), https://www.pacer.gov/documents/epa_feesched.pdf [http://perma.cc/72C8-GYEL].

[107] *Id*. (describing the fee schedule for reports generated by PACER).

years.[108] Retention rules can also play an important role in the analyses performed on records.[109] In the case of consumer arbitration, the CFPB should think carefully about how this relates to statutes of limitations and whether its collection should be retained for a longer period than record retention periods required of firms involved in the arbitration.

C. Responsibility for Privacy Protection and Required Disclosure Format

Identifying the privacy considerations is only one of the first steps in analyzing the benefits and burdens of an arbitral publication system. Two more important privacy-related steps remain. First, privacy considerations must be applied to the data. Second, the data must be transmitted in a manner that reflects the privacy considerations.

Responsibility for privacy protection under an archival arbitration scheme might fall either on financial institutions and arbitrators, or the CFPB itself. Determining who is required to apply the privacy principles will affect both the compliance costs and the usefulness of the data.[110] In deciding which entities should bear this responsibility, the CFPB must consider a covered person's technological sophistication and privacy expertise. The CFPB must also consider other subjective factors that may affect the data.

For example, two financial institutions that both have a thorough understanding of privacy practices may redact different amounts and types of data based on their institutional concerns. While one may under-redact if it is more concerned about the CFPB's scrutiny than potential harms to a consumer's privacy, the other may have aggressive privacy practices that cause it to over-redact.

---

[108] CAL. CIV. PROC. CODE § 1281.96(a) (West 2007) ("[A]ny arbitration company that administers or is otherwise involved in, a consumer arbitration shall . . . make available to the public . . . [arbitration information] within the preceding five years . . . .").

[109] *See* Lipner, *supra* note 77, at 57 ("The Article studies FINRA arbitrations in such cases and reveals that customer complaints regarding claims that later settled are being expunged at a rate of 93.7%, often in perfunctory ex parte proceedings where the complainant has agreed to not oppose the application as part of the settlement.").

[110] OUTLINE OF PROPOSALS, *supra* note 1, at 24 (discussing how privacy principles affect "cost of compliance with existing consumer finance and other laws and other costs due to entities attempting to minimize any such additional class litigation exposure in the future").

Similarly, if arbitrators are required to redact, and do not have a sophisticated understanding of privacy considerations, then the data transmitted may be over- or under-redacted. In either of these examples, the data submitted will be inconsistent, which will likely impair the usefulness of the data and potentially undermine the overall value of the arbitral publication system. Further, if financial institutions or arbitrators are required to apply the privacy principles, the CFPB must still develop processes to ensure that consumer privacy is protected in the event that a covered entity incorrectly submits data that should have been redacted. A privacy standard that avoids qualitative assessments (e.g., an enumerated list of proscribed identifiers) sidesteps these issues, but as discussed in Part II.B, also increases the risk of consumer harm.

These concerns could be avoided if the CFPB assumes responsibility for protecting privacy. However, this option raises an additional set of concerns. While the CFPB taking responsibility for privacy concerns would impose the lowest compliance costs on the industry, but it might take longer for the CFPB to scrub records than for those with knowledge of the facts and circumstances of individual cases. This may frustrate the goal of empowering the public with an arbitral publication system. Delays may also create an information asymmetry, where financial institutions know the current state of arbitration liability, but consumers and their representatives are relying on stale data. If the CFPB assumes responsibility, it might also obtain a substantial amount of consumer PII, increasing the magnitude and complexity of the CFPB's own internal privacy operations. Further, assuming responsibility may motivate the CFPB to not articulate the proposed privacy standard clearly. And if the proposed privacy standard is not sufficiently clear, consumer and industry groups may not be able to comment effectively on the value of the proposed arbitral publication system in the near term, and may not be able to judge the reliability of the system in the long term, as the privacy algorithm could evolve without further notice and comment.

Regardless of who is responsible for protecting privacy, the format of the data itself also plays a role in the burdens and benefits of the arbitral publication system. The more structured a dataset is, the more useful it is to data users. But the more sophisticated a data reporting structure is, the higher the implementation and ongoing costs tend to be. While financial institutions often possess, or are able to acquire, data processing and transmission tools, these tools come at a cost. And although financial institutions can spread the costs among a large number of transactions, it is unlikely an arbitral publication

system with the complexity of the SEC's EDGAR system would "impose minimal costs," as the CFPB asserts in its SBREFA materials.[111] On the other hand, if a less sophisticated, and less expensive, system is chosen, then data quality and usability will be adversely affected. For example, if arbitrators are required to transmit the data, they would likely have an easy time transmitting PDFs, but that format lacks structure and would significantly increase the time needed to process and disclose the data.

In sum, beyond merely identifying its privacy considerations, the CFPB must carefully consider how privacy requirements will be applied to the arbitration data, and how the data will be transmitted in a manner that reflects those privacy considerations. Again, these plans must be disclosed as a part of the CFPB's proposal, in order to give covered institutions and the public an opportunity to fully weigh the costs and benefits of an arbitration archive.

IV. Closing Thoughts

The CFPB's potential proposed rulemaking on consumer financial arbitration agreements is poised to require collecting data on consumer financial arbitral claims and awards, and possibly disseminating it to the public, potentially covering a wide range of business practices and contexts.[112] This article considers how data collection goals and practical considerations might inform the design of the data collection.

The CFPB must take the critical step to define its privacy considerations, particularly relating to consumers, and determine who is responsible for maintaining privacy standards. This step is particularly critical for products that may be impacted by other privacy standards, such as HIPAA, based on their relationship to products in industries with high privacy standards. Our analysis also considers the tradeoffs between comparatively freeform data collection and structured data collection that embeds categorization and classification into the data submission process. Considering the extensive use of hand-coding in previous studies of arbitration and the large amount of

---

[111] *Contra id.* at 20 ("This aspect of the proposal under consideration would not require changes to be made to the text of companies' arbitration agreements, alter the conduct of arbitration proceedings, or impose requirements on the content of written awards and, as discussed below in part V, would impose minimal costs on covered entities.").

[112] *See generally id.* (discussing the effects of the proposed rulemaking).

skilled labor hand-coding requires, ex-ante standardization may be the most efficient way to follow market trends. Other fundamental design considerations are the extent to which data is processed, and the extent to which data processing is financed through the CFPB's budget or usage fees. Regardless of the data collection and dissemination approach selected, implementing a collection that is consistent, durable, and useable for research and analysis requires ongoing investment of time and resources.