

6-29-2022

Have Disclosures Kept Up with the Big Data Revolution? An Empirical Test

Uri Benoliel

College of Law & Business, urib@clb.ac.il

Follow this and additional works at: <https://lawdigitalcommons.bc.edu/bclr>



Part of the [Business Organizations Law Commons](#), [Consumer Protection Law Commons](#), and the [Science and Technology Law Commons](#)

Recommended Citation

Uri Benoliel, *Have Disclosures Kept Up with the Big Data Revolution? An Empirical Test*, 63 B.C. L. Rev. 1913 (2022), <https://lawdigitalcommons.bc.edu/bclr/vol63/iss6/2>

This Article is brought to you for free and open access by the Law Journals at Digital Commons @ Boston College Law School. It has been accepted for inclusion in Boston College Law Review by an authorized editor of Digital Commons @ Boston College Law School. For more information, please contact abraham.bauer@bc.edu.

HAVE DISCLOSURES KEPT UP WITH THE BIG DATA REVOLUTION? AN EMPIRICAL TEST

URI BENOLIEL

INTRODUCTION	1914
I. THEORETICAL BACKGROUND.....	1918
<i>A. The Existing Literature on Human Readability</i>	1918
<i>B. Machine-Readable Disclosures</i>	1920
1. Characteristics of Machine-Readable Disclosures.....	1921
2. Social Benefits of Machine-Readable Disclosures	1924
II. THE EMPIRICAL TEST.....	1929
<i>A. The Quick Service Restaurant Franchise Industry</i>	1930
<i>B. Sample</i>	1931
<i>C. Methodology</i>	1932
1. The Franchisor Balance Sheet—An Overview	1932
2. Why Test Balance Sheet Disclosure?	1934
3. How Was Machine Readability Tested?.....	1936
<i>D. Results</i>	1937
1. Most Balance Sheets Are in a Digital Format That Allows for Some Level of Data Searching and Extraction	1937
2. Balance Sheets Do Not Have a Unified Structure That Would Facilitate <i>Systematic</i> Computerized Data Extraction.....	1938
3. Balance Sheets Often Include a Non-Standardized Taxonomy That Impedes Machine Reading.....	1939
4. Disclosures Do Not Employ Unique, Non-Duplicative Data Identifiers.....	1940
III. NORMATIVE DISCUSSION AND IMPLICATIONS.....	1940
CONCLUSION	1944
APPENDIX A: FRANCHISES STUDIED	1946

HAVE DISCLOSURES KEPT UP WITH THE BIG DATA REVOLUTION? AN EMPIRICAL TEST

URI BENOLIEL *

Abstract: Given the significant social benefits of the big data revolution, an important empirical legal question arises: are government-mandated disclosures designed in a way that allows society to harness the power of the big data that they include? Mandated disclosures normally include an overwhelming volume of data that can be difficult to read and understand for the average individual consumer. If, however, the voluminous data included in the disclosures is machine-readable, such that it can be automatically extracted and processed by computers, disclosures might actually assist consumers in making better-informed buying decisions. Although legal scholars have extensively studied the level of human readability of disclosures, they have yet to study their machine readability. This Article aims to fill this research gap. Using the important U.S. quick-service (fast food) restaurant franchise industry as a case study, this Article examines whether disclosure documents, provided by franchisors to prospective franchisees, have the features of machine-readable data. It specifically tests whether disclosures are provided in an adequate digital format, and include unique data identifiers, structured format, and standardized taxonomy, which can be easily read and processed by computers. The sample of this study includes the financial balance sheets disclosed by one hundred dominant quick-service restaurant chains, including Subway, McDonald's, KFC, and Dunkin'. The disturbing empirical results of this study indicate that franchise disclosures are normally non-machine readable. Given these results, this Article presents concrete recommendations to policy-makers on how to assure that disclosures in all industries keep up with the big data revolution.

INTRODUCTION

The big data revolution has greatly benefited society.¹ For example, in the health care industry, big data assisted scientists in the search for COVID-19

© 2022, Uri Benoiel. All rights reserved.

* Faculty of Law, College of Law & Business. J.S.D. (UC Berkeley); LL.M. (Columbia University). I thank Yonathan Arbel, Oren Bar-Gill, Samuel Becher, Christopher Drahozal, and Roy Shapira for excellent comments on a previous version; the College of Law & Business for generous financial support; and the participants at the 17th Annual Asian Law and Economics Association Conference.

¹ For the purposes of this article, big data is defined as “data sets, typically consisting of billions or trillions of records, that are so vast and complex that they require new and powerful computational resources to process.” *Big Data*, DICTIONARY.COM, <https://www.dictionary.com/browse/big-data> [<https://perma.cc/E4YZ-J2S3>]. For seminal books explaining how big data has revolutionized society, see, for example, VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, *BIG DATA: A REVOLUTION*

vaccines.² In the retail sector, companies such as Amazon.com, Inc. and Walmart Inc. use big data to predict customer preferences.³ In the entertainment industry, companies such as Netflix, Inc. use big data to design innovative content and improve user experience.⁴ In finance, the U.S. Securities and Exchange Commission (SEC) has been using big data to detect insider trading.⁵ Likewise, various entities are using big data in the banking industry to prevent money laundering.⁶ In the transportation safety arena, many enforcement agencies have been using big data to predict the location of future motor vehicle accidents.⁷ The legal tech industry, too, is harnessing the power of big data. To illustrate, Lex Machina, a legal analytics platform, provides strategic analysis of the behavior of judges and lawyers based on “millions of pages of

THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK (2013); BERNARD MARR, *BIG DATA IN PRACTICE: HOW 45 SUCCESSFUL COMPANIES USED BIG DATA ANALYTICS TO DELIVER EXTRAORDINARY RESULTS* (2016); PHIL SIMON, *TOO BIG TO IGNORE: THE BUSINESS CASE FOR BIG DATA* (2013). For further literature, see Joseph A. Tomain, *Big Data and the Fourth Estate: Protecting the Development of News Media Monitoring Databases*, 12 J. BUS. & TECH. L. 53, 53 (2016) (“‘Big data’ is revolutionizing our lives and society.”); Joseph A. Tomain, *Online Privacy and the First Amendment: An Opt-In Approach to Data Processing*, 83 U. CIN. L. REV. 1, 14 (2014) (“[T]he Big Data era has fundamentally changed our society in exceptional ways.”); Stephen R. Mitroff & Benjamin Sharpe, *Using Big Data to Solve Real Problems Through Academic and Industry Partnerships*, 18 CURRENT OP. BEHAV. SCIS. 91, 91 (2017) (“Big data has revolutionized a number of industries . . .”).

² Ken Kaplan, *Coronavirus Gives AI and Big Data Chance to Shine*, FORECAST BY NUTANIX (Mar. 19, 2020), <https://www.nutanix.com/theforecastbynutanix/technology/ai-and-big-data-fight-coronavirus> [<https://perma.cc/45DA-TX2V>] (stating that big data and artificial intelligence technologies are “helping scientists find a vaccine” for COVID-19); *The Importance of Big Data in COVID-19*, CEO VIEWS, <https://theceoviews.com/the-importance-of-big-data-in-covid-19/> [<https://perma.cc/928D-3XGV>] (“Big data has also been combined with deep learning and bioinformatics to streamline the condition by developing the vaccine against the virus.”).

³ See, e.g., Bernard Marr, *Amazon: Using Big Data to Understand Customers*, BERNARD MARR & CO. (July 23, 2021), <https://bernardmarr.com/amazon-using-big-data-to-understand-customers/> [<https://perma.cc/5KK6-N3GX>]; Bernard Marr, *Walmart: Big Data Analytics at the World’s Biggest Retailer*, BERNARD MARR & CO. (July 23, 2021), <https://bernardmarr.com/walmart-big-data-analytics-at-the-worlds-biggest-retailer/> [<https://perma.cc/9BUD-JWUN>].

⁴ See, e.g., Gabrielle Sadeh, *How Netflix Uses Big Data to Create Content and Enhance User Experience*, CLICKZ (Mar. 20, 2019), <https://www.clickz.com/how-netflix-uses-big-data-content/228201/> [<https://perma.cc/T4H3-H8U3>].

⁵ See *Here’s How the SEC Is Using Big Data to Catch Insider Trading*, FORTUNE (Nov. 1, 2016), <https://fortune.com/2016/11/01/sec-big-data-insider-trading/> [<https://perma.cc/G7VP-T7TB>].

⁶ See, e.g., Jackie Wheeler, *How Big Data Can Improve AML Compliance Processes*, JUMIO (Apr. 28, 2021), <https://www.jumio.com/big-data-aml-compliance/> [<https://perma.cc/CJN9-P8F2>] (describing the ways in which big data can add flexibility to anti-money laundering procedures); Venkatesan Marimuthu, *Applicability of Big Data for Effective Anti-money Laundering*, WIPRO (May 2017), <https://www.wipro.com/en-JP/blogs/venkatesan-marimuthu/applicability-of-big-data-for-effective-anti-money-laundering/> [<https://perma.cc/FDD4-XT4T>].

⁷ Jenni Bergal, *Troopers Use ‘Big Data’ to Predict Crash Sites*, PEW: STATELINE (Feb. 9, 2017), <https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2017/02/09/troopers-use-big-data-to-predict-crash-sites> [<https://perma.cc/8RH6-FBDD>].

legal” data.⁸ Similarly, LawGeex, a contract review automation platform, reviews and remediates incoming contracts based on a training dataset of “half a million contracts.”⁹

Considering the significant benefits of big data, an important legal empirical question arises: are government-mandated disclosures designed in a way that allows society to harness the power of the big data that they include?¹⁰ Firms are mandated to publish disclosures pursuant to numerous disclosure laws, which are among the most popular types of regulation in American law.¹¹ Such laws typically require the “discloser” to give the “disclosee” (for example, a consumer) a precontractual “disclosure” that contains valuable data about the features, costs, and benefits of a product or service that the discloser provides.¹² These disclosures cover a wide range of products and services such as franchises, securities, employee benefit plans, electronic fund transfers, product warranties, health insurance, credit cards, mortgages, savings and checking accounts.¹³ The

⁸ *New Comparator Apps from Lex Machina Empower Lawyers to Instantly Compare Courts, Judges, and Law Firms*, LEX MACHINA (Sept. 20, 2016), <https://lexmachina.com/media/press/new-comparator-apps/> [<https://perma.cc/B7XD-T79D>].

⁹ Meghan Han, “My Other Lawyer Is a Robot”—LawGeex Automates Contract Review, MEDIUM (Mar. 21, 2018), <https://medium.com/syncedreview/my-other-lawyer-is-a-robot-lawgeex-automates-contract-review-eef4e2247114> [<https://perma.cc/83ZA-SHZ2>].

¹⁰ For a prior empirical study of human-readable disclosures, see generally Uri Benoliel & Xu (Vivian) Zheng, *Are Disclosures Readable? An Empirical Test*, 70 ALA. L. REV. (2018).

¹¹ Omri Ben-Shahar & Carl E. Schneider, *The Futility of Cost-Benefit Analysis in Financial Disclosure Regulation*, 43 J. LEGAL STUD. S253, S253 (2014) (“Mandated disclosure is one of the most common regulatory techniques in American law.”); Florencia Marotta-Wurgler, *Does Contract Disclosure Matter?*, 168 J. INSTITUTIONAL & THEORETICAL ECON. 94, 94 (2012) (“Disclosure has long been the preferred regulatory approach to prevent one-sided standard-form contract terms”); Matthew A. Edwards, *The Virtue of Mandatory Disclosure*, 28 NOTRE DAME J.L., ETHICS & PUB. POL’Y 47, 47 (2014) (“During the past fifty years mandatory disclosure has emerged as a dominant method of legal regulation in the United States.” (first citing ARCHON FUNG, MARY GRAHAM & DAVID WEIL, *FULL DISCLOSURE: THE PERILS AND PROMISE OF TRANSPARENCY* 20–24 (2007); then citing MARY GRAHAM, *DEMOCRACY BY DISCLOSURE: THE RISE OF TECHNOPOPULISM* 4 (2002); and then citing Cass R. Sunstein, *Informational Regulation and Informational Standing: Akins and Beyond*, 147 U. PA. L. REV. 613, 619 (1999)); Daniel E. Ho, *Fudging the Nudge: Information Disclosure and Restaurant Grading*, 122 YALE L.J. 574, 577 (2012) (stating that “disclosure is a mainstay of the regulatory toolkit”); see also Erin Bernstein, *The Upside of Abortion Disclosure Laws*, 24 STAN. L. & POL’Y REV. 171, 190 (2013) (“Indeed, disclosure laws are commonplace” (citing Omri Ben-Shahar & Carl E. Schneider, *The Failure of Mandated Disclosure*, 159 U. PA. L. REV. 647 (2011))); Paula J. Dalley, *The Use and Misuse of Disclosure as a Regulatory System*, 34 FLA. ST. U. L. REV. 1089, 1092 (2007) (“There are dozens, possibly hundreds, of regulatory schemes that use disclosure in whole or in part to accomplish their purposes.”).

¹² See, e.g., Dalley, *supra* note 11, at 1093–94 (noting that the Securities Act of 1933 obligates issuers to disclose information before selling a security to the public, and that pursuant to the Securities Exchange Act of 1934, publicly traded companies must disclose certain information to shareholders).

¹³ See, e.g., The Franchise Rule, 16 C.F.R. pts. 436–37 (2022) (requiring disclosures to allow potential franchisees to weigh the risks and benefits of the offered franchise before making a purchasing decision); General Rules and Regulations, Securities Act of 1933, 17 C.F.R. pt. 230 (2022) (requiring

disclosures often include an overwhelming volume of data about the products or services, which individual consumers can hardly read or understand.¹⁴ If, however, the voluminous data included in mandated disclosures is machine-readable—meaning computers can extract and process it automatically—data aggregators can efficiently gather the extracted data into analytical reports that consumers can then use to make better-informed buying decisions.¹⁵ Therefore, in light of the potential benefits of machine-readable disclosures, this Article empirically examines whether disclosures are indeed machine-readable.¹⁶

Namely, using the important U.S. quick-service (fast food) restaurant franchise industry as a case study, this Article examines whether franchise disclosure documents have the features of machine-readable data.¹⁷ For that purpose, the Article investigates the content and structure of a sample of financial balance sheets disclosed by one hundred leading restaurant chains, including Subway, McDonald's, KFC, and Dunkin'.¹⁸

This Article is structured as follows: Part I provides the theoretical context for the empirical test of this study.¹⁹ It first presents the existing literature on the human readability of disclosures and the contribution of this Article to this literature.²⁰ Thereafter, it explains what machine-readable disclosures are, and presents the potential social benefits of these disclosures.²¹ Part II presents data and discusses the methodology for empirically testing the machine reada-

disclosures to ensure that buyers of securities receive complete and accurate information before they invest in securities); Employee Retirement Income Security Act of 1974, 29 U.S.C. § 1001(b) (aiming to protect the interests of employee benefit plan participants by requiring plan sponsors to disclose financial data to plan participants and beneficiaries); 12 C.F.R. § 205.1(b) (2022) (requiring financial institutions to disclose the terms of electronic fund transfer activities, in order to protect the protect consumers engaging in these activities); Magnuson-Moss Warranty—Federal Trade Commission Improvement Act, 15 U.S.C. §§ 2301–2312 (aiming to ensure that consumers get an opportunity to compare warranty coverage before buying products); The HIPAA Privacy Rule, 45 C.F.R. § 164.520(a)(1) (2022) (providing individuals a right to be informed of the privacy practices of their health plans); Truth in Lending Act, 15 U.S.C. § 1601(a) (requiring lenders to disclose to potential borrowers data about the cost of a loan before the formation of a loan agreement). On the various types of disclosures see generally OMRI BEN-SHAHAR & CARL E. SCHNEIDER, MORE THAN YOU WANTED TO KNOW: THE FAILURE OF MANDATED DISCLOSURE 4 (2014).

¹⁴ See BEN-SHAHAR & SCHNEIDER, *supra* note 13, at 79–80 (“Many people cannot read many disclosures because they are not literate or numerate enough to decipher them with reasonable effort.”); Oren Bar-Gill, *Defending (Smart) Disclosure: A Comment on More Than You Wanted to Know*, 11 JERUSALEM REV. LEGAL STUD. 75, 75 (2015) (reviewing BEN-SHAHAR & SCHNEIDER, *supra* note 13) (“Consumers are not going to read the long, detailed disclosure statements that so many laws and regulations mandate.”).

¹⁵ See *infra* Section I.B.

¹⁶ See *infra* Section II.C.

¹⁷ See *infra* Part II.

¹⁸ See *infra* Sections II.B and II.C.

¹⁹ See *infra* notes 24–91 and accompanying text.

²⁰ See *infra* notes 27–42 and accompanying text.

²¹ See *infra* notes 43–91 and accompanying text.

bility of disclosures.²² Part III discusses the normative implications of the empirical results.²³

I. THEORETICAL BACKGROUND

This Part offers the theoretical context that grounds the empirical test of the study presented herein.²⁴ Section A provides an overview of the existing scholarship regarding disclosure human readability, and also explains the unique contribution that this Article makes to the conversation.²⁵ Section B covers the main characteristics of machine-readable disclosures and describes some of their various societal benefits.²⁶

A. *The Existing Literature on Human Readability*

The readability of disclosures—that is, the ease with which the written text of disclosures can be read and understood—has already been studied by legal scholars and institutions.²⁷ These studies have focused on the non-readability of disclosures by *humans*, normally consumers, as opposed to computers.²⁸ For example, the seminal disclosure law book *More Than You Wanted to Know: The Failure of Mandated Disclosure* explains that consumers often cannot read disclosures because the reading level required to understand a disclosure is unexpectedly high.²⁹ Similarly, another scholar who focuses on the non-readability of disclosures by humans, explained that “[c]onsumers are not going to read the long, detailed disclosure statements that so many laws and regulations mandate.”³⁰

The readability of disclosures to humans, as opposed to machines, has also been the subject of systematic empirical studies. To illustrate, a study con-

²² See *infra* notes 96–179 and accompanying text.

²³ See *infra* notes 180–208 and accompanying text.

²⁴ See *infra* notes 27–91 and accompanying text.

²⁵ See *infra* notes 27–43 and accompanying text.

²⁶ See *infra* notes 43–91 and accompanying text.

²⁷ See, e.g., BEN-SHAHAR & SCHNEIDER, *supra* note 13, at 79–80; Benoliel & Zheng, *supra* note 10, at 237, 247–51; O. Seizov, A.J. Wulf & J. Luzak, *The Transparent Trap: A Multidisciplinary Perspective on the Design of Transparent Online Disclosures in the EU*, 42 J. CONSUMER POL’Y 149, 161 (2019); U.S. GOV’T ACCOUNTABILITY OFF., GAO-06-929, CREDIT CARDS: INCREASED COMPLEXITY IN RATES AND FEES HEIGHTENS NEED FOR MORE EFFECTIVE DISCLOSURES TO CONSUMERS 3, 11–12 (2006) [hereinafter CREDIT CARDS: INCREASED COMPLEXITY]; U.S. GOV’T ACCOUNTABILITY OFF., GAO-14-92, PRIVATE PENSIONS: CLARITY OF REQUIRED REPORTS AND DISCLOSURES COULD BE IMPROVED 2–3, 45 (2013) [hereinafter PRIVATE PENSIONS].

²⁸ See, e.g., Benoliel & Zheng, *supra* note 10, at 239, 257 (evaluating the extent to which disclosures are readable by humans); CREDIT CARDS: INCREASED COMPLEXITY, *supra* note 27, at 6 (evaluating the human readability of disclosures made to credit card holders).

²⁹ BEN-SHAHAR & SCHNEIDER, *supra* note 13, at 79–80.

³⁰ Bar-Gill, *supra* note 14, at 75.

ducted by the U.S. Government Accountability Office (GAO), tested the human readability of the written offers that credit card issuers provide to their consumers, which disclose important information about the interest rates, fees, and related charges associated with using such cards.³¹ To assess readability, the study used linguistic formulas that calculated the average syllable-count in each word and the average word-count in each sentence, and ultimately found that the written offers “were too complicated for many consumers to understand.”³² Another study, also conducted by the GAO, assessed the human readability of sixteen “model notices,” namely templates created by the Department of Labor and the Internal Revenue Service to help pension-plan sponsors draft disclosures to plan participants.³³ Using human readability formulas, the study found that the readability of the model notices examined in the study “might present a challenge for participants.”³⁴

In the context of franchising—the focus of the empirical test in this Article—researchers have also studied the readability of disclosures to humans (for example, potential franchisees). For example, a franchise law scholar explained that “[t]he required [franchise] disclosure document can be quite a lengthy read for prospective franchisees. The sheer number of pages in the typical disclosure can be intimidating and is probably a reason so many franchisees skip the important step of carefully evaluating the material information contained inside the document.”³⁵ Similarly, franchise law attorneys have uncovered that the amount of data contained in a typical franchise disclosure document has increased in recent years.³⁶ They also suggest that “[s]uch an increase in information makes the document impenetrable and intimidating to most potential franchisees.”³⁷ Likewise, and from an empirical perspective, my study with Professor Xu (Vivian) Zheng empirically examined the human

³¹ CREDIT CARDS: INCREASED COMPLEXITY, *supra* note 27, at 3, 11–12.

³² *Id.* at 6, 37, 83.

³³ PRIVATE PENSIONS, *supra* note 27, at 2–3, 45.

³⁴ *Id.* at 36. Notably, other empirical studies examined the human readability of consumer contracts, as opposed to precontractual disclosures. *See, e.g.*, Florencia Marotta-Wurgler & Robert Taylor, *Set in Stone? Change and Innovation in Consumer Standard-Form Contracts*, 88 N.Y.U. L. REV. 240, 253–54 (2013); Uri Benoliel & Shmuel I. Becher, *The Duty to Read the Unreadable*, 60 B.C. L. REV. 2255, 2271–77 (2019).

³⁵ Robert W. Emerson, *Transparency in Franchising*, 2021 COLUM. BUS. L. REV. 172, 204.

³⁶ Eric H. Karp & Ari N. Stern, *A Proposal for a Mandatory Summary Franchise Disclosure Document*, 35 FRANCHISE L.J. 541, 543 (2016) (stating that a development in FDDs “has been the material and marked increase in the sheer amount of information contained in a typical franchise disclosure document”).

³⁷ *Id.*; *see also* Rochelle Spandorf, *Reading the FDD: The Argument Against Simplification*, FRANCHISING.COM (May 4, 2021), https://www.franchising.com/articles/reading_the_fdd_the_argument_against_simplification.html [<https://perma.cc/VN3Q-Y9K4>] (“[T]he FDD has grown so complicated and voluminous that it is ‘intimidating’ to the average prospective franchisee . . .”).

readability of disclosures in the franchise industry.³⁸ Using the Gunning Fog Index human readability formula, which is based on the average number of words in each sentence and the percentage of complex words,³⁹ the study indicated that disclosures are normally unreadable by potential franchisees.⁴⁰ The study particularly indicated that prospective “franchisees need, on average, more than twenty years of education” to understand a franchise disclosure document on the first reading, while most franchisees only have fourteen years of education.⁴¹

Notwithstanding the paramount importance of disclosures in U.S. law, there has been surprisingly little empirical study of their *machine*-readability, as opposed to their human-readability. One exceptional study, published in 2014, documented, in general terms, that hydraulic fracturing chemical disclosure data in many U.S. state disclosure websites are not machine-readable.⁴² This Article aims to contribute to this scant empirical research. It attempts to systematically test the machine-readability of disclosures that sellers provide to their potential buyers. Particularly, the test in this Article is based on a thorough investigation of the content and structure of real-world disclosures. The Article specifically examines whether disclosures exist in an adequate digital format, and whether they include unique data identifiers, structured format, and standardized taxonomy that can be easily read and processed by computers.

B. Machine-Readable Disclosures

This Section begins by presenting the major characteristics of machine-readable disclosures. These features underlie the following empirical test of this study. The Section then presents the potential social benefits of machine-readable disclosures. These benefits highlight the importance of empirically examining whether disclosures are machine-readable, as carried out in this study.

³⁸ Benoiel & Zheng, *supra* note 10, at 247–51.

³⁹ Aymen Ajina, Mhamed Laouti & Badreddine Msolli, *Guiding Through the Fog: Does Annual Report Readability Reveal Earnings Management?*, 38 RSCH. INT’L BUS. & FIN. 509, 513 (2016); J. Efrim Boritz, Louise Hayes & Lev M. Timoshenko, *Determinants of the Readability of SOX 404 Reports*, 13 J. EMERGING TECHS. ACCT. 145, 154 n.7 (2016). This Gunning Fog Index is one of the most popular methods of measuring text readability and is based on these two premises: “(1) more words per sentence increase the probability that the text will be harder to read; (2) more syllables per word increase the probability that the text will be harder to read.” Benoiel & Zheng, *supra* note 10, at 249–50 (first citing Judith Bogert, *In Defense of the Fog Index*, BULL. ASS’N FOR BUS. COMM’N, June 1985, at 9, 10; then citing Bortiz et al., *supra*, at 146; and then citing Feng Li, *Annual Report Readability, Current Earnings, and Earnings Persistence*, 45 J. ACCT. & ECON. 221, 222 (2008)).

⁴⁰ Benoiel & Zheng, *supra* note 10, at 257.

⁴¹ *Id.*

⁴² Matthew McFeeley, *Falling Through the Cracks: Public Information and the Patchwork of Hydraulic Fracturing Disclosure Laws*, 38 VT. L. REV. 849, 864–70 (2014).

1. Characteristics of Machine-Readable Disclosures

Machine-readable disclosures are documents in a format that can be easily and accurately processed by computers.⁴³ In order to be seamlessly and accurately processed by computers, machine-readable disclosures must be structured.⁴⁴ Structured disclosures should fulfill the following four major conditions:

- (1) They should be in a *digital format* that can be easily searched via computer algorithms.⁴⁵
- (2) Structured disclosures, in each industry, must all be designed in a *standardized structure*.⁴⁶ This uniform structure assists computers to extract into database applications the data systematically from the various disclosures provided by the different disclosers in the industry.⁴⁷

⁴³ Cf. 44 U.S.C. § 3502(18) (stating that “the term ‘machine-readable’, when used with respect to data, means data in a format that can be easily processed by a computer without human intervention while ensuring no semantic meaning is lost”); *Machine Readable*, OPEN DATA HANDBOOK, <https://opendatahandbook.org/glossary/en/terms/machine-readable/> [<https://perma.cc/LA3F-J7VE>] (stating machine readable data is in a format that “can be automatically read and processed by a computer”).

⁴⁴ Cf. *Machine Readable*, *supra* note 43 (“Machine-readable data must be structured data.”).

⁴⁵ Cf. *Big Data Analysis by Combining Structured & Unstructured Data*, MARUTI TECHLABS, <https://marutitech.com/big-data-analysis-structured-unstructured-data/> [<https://perma.cc/ZNL2-UXXG>] (explaining that “structured data is relatively easy to search”); Mark Smallcombe, *Structured vs Unstructured Data: 5 Key Differences*, INTEGRATE.IO (Jan. 03, 2022), <https://www.integrate.io/blog/structured-vs-unstructured-data-key-differences/> [<https://perma.cc/VH9Y-FP5A>] (“Structured data [are] clearly defined and searchable types of data . . .”).

⁴⁶ Cf. Xiaoqiu Le, Chenyu Mao, Yuanbiao He, Changlei Fu & Liyuan Xu, *Dpaper: An Authoring Tool for Extractable Digital Papers*, 1 J. DATA & INFO. SCI. 86, 88 (2016) (“An extractable document should . . . [b]e structured . . .”); *Understand How Structured Data Works*, GOOGLE, <https://developers.google.com/search/docs/advanced/structured-data/intro-structured-data> [<https://perma.cc/2CQN-236L>] (“Structured data is a standardized format for providing information about a page and classifying the page content . . .”); *What Is Structured Data?*, U.S. SEC. & EXCH. COMM’N, <https://www.sec.gov/structureddata/what-is-structured-data> [<https://perma.cc/UQP7-4W9K>] (Mar. 26, 2016) (“Structured data [are] data that [are] divided into standardized pieces . . .”).

⁴⁷ Cf. *Data Types: Structured vs. Unstructured Data*, ENTER. BIG DATA FRAMEWORK (Jan. 9, 2019), <https://www.bigdataframework.org/data-types-structured-vs-unstructured-data/> [<https://perma.cc/HZ7K-VZZ2>] (noting that when data are structured “it is possible to quickly aggregate data from various locations in the database”); *What Is Structured Data?*, *supra* note 46 (“Widely available software can be used to easily analyze vast amount of structured data without extensive and burdensome manual processing.”); see Memorandum from Cass R. Sunstein, Adm’r, Off. of Info. & Regul. Affs., to Heads of Exec. Dep’ts & Agencies 5 (Sept. 8, 2011), https://www.nist.gov/system/files/documents/ineap/Summit_Invitation_to_Agencies_FINAL.pdf [<https://perma.cc/QZ7V-CB4X>] (explaining that data in a machine-readable format “may be readily imported into spreadsheet and database applications”).

- (3) Structured disclosures in each industry must use *standardized taxonomy* elements that enable computers to detect and compare these elements easily among various disclosures.⁴⁸
- (4) Each data item in the disclosure must have a *unique* data identifier (i.e., unique title or symbol) that differentiates it from other data items in the same disclosure.⁴⁹ This unique identifier assists computers to locate a relevant data item easily within the disclosure, and prevents a computer from conflating different data items with duplicate data identifiers.⁵⁰

To provide a basic example of structured data, sales reports that various companies provide are structured to the extent that: 1) the sales reports are provided in a digital Excel format that can be easily searched and extracted by computers; 2) the report of each company uses a uniform structure of four columns that represent the customer's first name, the customer's last name, the customer's amount of purchase in U.S. dollars, and the date of purchase; 3) each of these columns has the following uniform titles: "first name," "last name," "amount," and "date;" and 4) these titles are unique and non-duplicative, in each report.⁵¹

Conversely, a basic example of unstructured data, which cannot be easily and accurately processed by computers, is scanned images of the body of casu-

⁴⁸ Memorandum from Cass R. Sunstein, Adm'r, Off. of Info. & Regul. Affs., to Heads of Exec. Dep'ts & Agencies, *supra* note 47, at 5 ("Standardized vocabularies and formats allow for meaningful comparisons and other analyses across datasets."); Stephen Camilleri, *XBRL and ESF Reporting*, DELOITTE, <https://www2.deloitte.com/mt/en/pages/strategy-operations/articles/mt-consulting-article-xbrl-and-esef-reporting.html> [<https://perma.cc/89EV-NZAW>] (defining taxonomy as "a dictionary of elements, or tags, that represent the reporting fields/concepts required for specific statement filings").

⁴⁹ Cf. Matt G. Southern, *Google: Structured Data Should Be Unique to Each Page*, SEARCH ENGINE J. (Feb. 23, 2018), <https://www.searchenginejournal.com/google-structured-data-unique-page/239507/#close> [<https://perma.cc/R6WL-RA6E>] ("If a page needs to contain structured data at all, then it should be unique to that specific page."); Lindsey Nelson, *Structured Data*, HUCKABUY, <https://huckabuy.com/technical-seo/structured-data/> [<https://perma.cc/MMX3-2E2P>] (Dec. 2, 2021) ("Each structured data object has a unique set of recommended and required properties . . .").

⁵⁰ Cf. Irene Dobik, *Guide to SEO First Aid: Errors to Fix Without a Specialist*, APPINSTITUTE (Sept. 11, 2017), <https://appinstitute.com/seo-first-aid/> [<https://perma.cc/AU5A-LFYX>] ("Unique titles prevent your pages from competing with each other and help avoid confusion among search engines.").

⁵¹ For a similar instance, see, for example, *Structured vs Unstructured Data: Compared and Explained*, ALTEXSOFT (Dec. 14, 2020), <https://www.altexsoft.com/blog/structured-unstructured-data/> [<https://perma.cc/M7J9-BTHM>] (presenting, as an example of a typical structured data, an "Excel spreadsheet that contains information about customers and purchases"); see also *Structured Data vs. Unstructured Data—What's the Difference?*, TREEHOUSE TECH. GRP., <https://treehouse.techgroup.com/structured-data-vs-unstructured-data-whats-the-difference/> [<https://perma.cc/KH2C-EE5Z>] ("Structured data is convenient to store, analyze and report. As an example, consider the sales report of a business, that shows sales revenue by region, by product, by month, and so on.").

ally-written email messages composed by various authors.⁵² These images are unstructured for three major reasons. First, the text in such images is often not easily or accurately searchable by a computer.⁵³ Before a computer can search such images, the images must be first converted, via an optical character recognition (OCR) method, to searchable text.⁵⁴ Moreover, the OCR conversion method is often inaccurate and the resulting text “contains errors that typically involve character exchanges.”⁵⁵ Second, the body of casual email messages normally consists of freeform text without standardized structure.⁵⁶ Lastly, the body of email messages composed by various authors normally does not systematically apply standardized taxonomy elements.⁵⁷

⁵² Mitchell Sloan, *Converting Unstructured Data in Documents to Structured Data*, ACODIS (Apr. 7, 2021), <https://www.acodis.io/blog/converting-unstructured-data-to-structured-data> [<https://perma.cc/MP2R-3WKG>] (listing examples of unstructured data, which include emails, reports, and multimedia content); Kimi Mahajan, *Processing Unstructured Data in 5 Easy Ways*, SRIJAN (Jan. 31, 2020), <https://www.srijan.net/resources/blog/processing-unstructured-data-in-5-easy-ways> [<https://perma.cc/9PSU-33G4>] (noting that unstructured data include scanned documents and emails).

⁵³ *Committee Begins Reconsideration of E-Signatures and Related Rules*, FLA. BAR NEWS (June 18, 2021), <https://www.floridabar.org/the-florida-bar-news/committee-begins-reconsideration-of-e-signatures-and-related-rules/> [<https://perma.cc/38CS-BF8T>] (“Scanned documents are not searchable”); *Working with PDF Files—An Ultimate Guide [2021]*, COPYLEAKS (Mar. 25, 2021), <https://copyleaks.com/blog/working-with-pdf-files-an-ultimate-guide-by-copyleaks-plagiarism-detector/> [<https://perma.cc/7VH9-9B3V>] (same). See generally *Machine Readable*, *supra* note 43 (“[S]cans (photographs) of text are not machine-readable”); Memorandum from Cass R. Sunstein, Adm’r, Off. of Info. & Regul. Affs., to Heads of Exec. Dep’ts & Agencies, *supra* note 47, at 5 (“[C]omputer files that are simply image reproductions of print disclosures . . . generally do not represent machine readable formats.”).

⁵⁴ *Digitize Your Sources (DIY)*, TUL. UNIV. LIBRS., https://libguides.tulane.edu/diy_digital/ocr [<https://perma.cc/488V-EGZV>] (Nov. 9, 2021) (“OCR (Optical Character Recognition) is a technology that enables you to convert different types of textual documents, such as scanned paper documents, PDF files or images captured by a digital camera into editable and searchable data.”).

⁵⁵ Guilherme Torresan Bazzo, Gustavo Acauan Lorentz, Danny Suarez Vargas & Viviane P. Moreira, *Assessing the Impact of OCR Errors in Information Retrieval*, in *ADVANCES IN INFORMATION RETRIEVAL: 42ND EUROPEAN CONFERENCE ON IR RESEARCH, ECIR 2020, LISBON, PORTUGAL, APRIL 14–17, 2020, PROCEEDINGS, PART II 102, 102* (Joemon M. Jose et al. eds., 2020).

⁵⁶ Pierre Dorion, *What Is Unstructured Data and How Is It Different from Structured Data in the Enterprise?*, TECHTARGET (Mar. 27, 2007), <https://searchstorage.techtarget.com/feature/What-is-unstructured-data-and-how-is-it-different-from-structured-data-in-the-enterprise> [<https://perma.cc/LQ3N-KY8M>] (“An email is considered unstructured data . . . the body of the message is really freeform text without any structure at all”); see also Memorandum from Cass R. Sunstein, Adm’r, Off. of Info. & Regul. Affs., to Heads of Exec. Dep’ts & Agencies, *supra* note 47, at 5 (“[C]omputer files . . . that contain only unstructured narrative text generally do not represent machine readable formats.”).

⁵⁷ Iggy Durant, *What Is the Best Way to Write an Email?*, PEEP STRATEGY (July 9, 2021), <https://peepstrategy.com/what-is-the-best-way-to-write-an-email/> [<https://perma.cc/ULZ4-7C2R>] (demonstrating that people generally do not subscribe to a standardized format when drafting the body of an email message and providing examples of various approaches to email drafting). Different email recipients and circumstances require different email styles, which militate against a standardized taxonomy. *Id.*

2. Social Benefits of Machine-Readable Disclosures

Machine-readable disclosures that computers can easily and accurately process have several important social advantages. To begin, these disclosures create significant value for data aggregators and their customers.⁵⁸ Data aggregators are organizations that collect voluminous data about specific products or services from many resources and present it to potential buyers in summary format (for example, comparison tables).⁵⁹ The aggregators frequently collect the data from government-mandated disclosures.⁶⁰ To illustrate, QSR.com publishes an annual online ranking of quick service franchises based on variables such as systemwide sales, restaurant unit count, and average restaurant sales per restaurant unit.⁶¹ QSR.com collects these variables from franchise mandated disclosures.⁶² Similarly, Creditcards.com compares numerous online credit card disclosure items, such as rates and fees among various credit card categories.⁶³

Machine-readable disclosures produce three major benefits to data aggregators which are presented below: a) cost saving; b) timesaving; and c) error-

⁵⁸ Michael Laermann, *Machine-Readable Disclosure and ESEF 2020—A New Era of Corporate Digital Reporting?*, SUSTAINABLE BRANDS (May 23, 2018), <https://sustainablebrands.com/read/marketing-and-comms/machine-readable-disclosure-and-esef-2020-a-new-era-of-corporate-digital-reporting> [<https://perma.cc/9BKF-7NRL>] (“The idea behind the machine-readable format is to make corporate reporting more accessible, comparable and transparent for . . . Big Data aggregators . . . who could feed their evaluation models with comparable data automatically . . .”); Letter from Campbell Pryde, President & CEO, XBRL US, Inc., to Brent J. Fields, Sec’y, U.S. Sec. & Exch. Comm’n, on Fund Retail Investor Experience and Disclosure, File No. S7-12-18 2 (Oct. 31, 2018), <https://www.sec.gov/comments/s7-12-18/s71218-4587748-176292.pdf> [<https://perma.cc/J6QL-35XT>] [hereinafter Letter on Fund Retail Investor Experience and Disclosure] (“[T]he benefits of computer-readable data . . . benefit retail investors, in the form of more timely and affordable data from commercial data aggregator . . . providers.”).

⁵⁹ For a definition of data aggregators, see DAVID LOSHIN, *BUSINESS INTELLIGENCE: THE SAVVY MANAGER’S GUIDE* 43 (2d ed. 2013) (“Data aggregators . . . are businesses that collect industry- or societywide data and enhance and reorganize that data as a way of providing value-added services to customers and subscribers.”); Craig S. Mullins, *Data Aggregation*, TECHTARGET, <https://searchsqlserver.techtargget.com/definition/data-aggregation> [<https://perma.cc/4S4R-DY85>] (June 2020) (“Data aggregators work by combining atomic data from multiple sources, processing the data for new insights and presenting the aggregate data in a summary view.”).

⁶⁰ BEN-SHAHAR & SCHNEIDER, *supra* note 13, at 187 (explaining that data aggregators “often collect the information that is in mandated disclosures”).

⁶¹ *Ranking the Top 50 Fast-Food Chains in America*, QSR MAG. (2020), <https://www.qsrmagazine.com/content/qsr50-2020-top-50-chart> [<https://perma.cc/6MDN-HKA2>].

⁶² *See id.* (“Filings from the . . . Franchise Disclosure Documents (FDD) . . . were amalgamated, reviewed, and analyzed to yield domestic sales, unit count, and average unit volume.”).

⁶³ *See, e.g.*, Jeanine Skowronski, *The Best Credit Cards of April 2022: Rewards, Top Offers & Reviews*, CREDITCARDS.COM, <https://www.creditcards.com/best-credit-cards/> [<https://perma.cc/TG5V-54VR>] (Apr. 4, 2022); *see also* BEN-SHAHAR & SCHNEIDER, *supra* note 13, at 187 (“Services like Creditcards.com compare the information about rates, fees, and much more that is in credit-card disclosures.”).

reduction. Subsection d discusses the value of disclosures to law enforcement and legal scholars.

a. The Cost-Saving Value of Machine-Readable Disclosures

When the data in disclosures is not machine-readable, it normally must be collected manually by human personnel.⁶⁴ This manual collection of data is costly since disclosures are often lengthy or numerous (or both). In the franchise industry, the typical disclosure document, released by each franchisor, includes several hundred pages of data.⁶⁵ Subway's disclosure, as an example, includes 676 pages.⁶⁶ Similarly, McDonald's disclosure is 399 pages long.⁶⁷ In addition, the total number of franchise disclosure documents, submitted by all franchisors each year, is significant. To illustrate, the number of effective annual disclosures publicly available on Wisconsin's Department of Financial Institutions website is more than 1,500.⁶⁸

When disclosures are machine-readable, however, data aggregators can automatically extract data, via computer software, from lengthy and numerous disclosures, and thereby save significant labor-intensive manual extraction costs.⁶⁹ The manual cost-saving achieved by the automatic extraction of data

⁶⁴ See Interactive Data to Improve Financial Reporting, Securities Act Release No. 9,002, Exchange Act Release No. 59,324, Trust Indenture Act Release No. 2,461, Investment Company Act Release No. 28,609, 74 Fed. Reg. 6,776, 6,803 (Feb. 10, 2009) ("Investors seeking broad financial coverage of companies must . . . spend considerable time manually collecting the data . . ."); Camilleri, *supra* note 48 (stating that when data is not machine-readable, regulators must "manually extract, tag, classify, rekey and integrate all the data from all submitting entities").

⁶⁵ See, e.g., *Franchise Disclosure Documents*, THE EDUCATED FRANCHISEE, <https://fdexchange.com/> [<https://perma.cc/B3U4-BNRV>] ("Franchise Disclosure Documents (FDD) provide hundreds of pages of information on franchise systems."); *FDD Guide: An Explanation by Popular Questions (Part 1)*, FRANCHISE DIRECT, <https://www.franchisedirect.com/information/introductiontofranchising/fddguideanexplanationbypopularquestionspart1/7/1091/> [<https://perma.cc/CLN7-SPD7>] (noting that an FDD "typically cover[s] hundreds of pages").

⁶⁶ See Dr.'s Assocs. LLC dba Subway, Franchise Disclosure Document (2021).

⁶⁷ See McDonald's USA, LLC, Franchise Disclosure Document (2021).

⁶⁸ *List of Active Registrations*, WIS. DEP'T OF FIN. INSTS., <https://www.wdfi.org/apps/franchiseefiling/activeFilings.aspx> [<https://perma.cc/WZV8-8V3R>].

⁶⁹ Letter on Fund Retail Investor Experience and Disclosure, *supra* note 58, at 2 (explaining that when financial disclosures are machine-readable they "can be processed automatically, which reduces the cost of data"); Letter from Campbell Pryde, President & CEO, XBRL US, Inc., to Sec'y, U.S. Sec. & Exch. Comm'n, on Investment Company Reporting Modernization, File No. S7-08-15 2 (Aug. 11, 2015), <https://www.sec.gov/comments/s7-08-15/s70815-288.pdf> [<https://perma.cc/LR5E-GPC5>] [hereinafter Letter on Investment Company Reporting Modernization] (noting that automatic extraction of structured data leads to "cost savings, because labor-intensive resources to rekey, review and validate data extracted from HTML or text files are no longer needed"); Allison Herren Lee, Comm'r, U.S. Sec. & Exch. Comm'n, *The Promise of Structured Data: True Modernization of Disclosure Effectiveness* (Nov. 17, 2020), <https://www.sec.gov/news/speech/lee-structured-data-2020-11-17> [<https://perma.cc/656E-2BRT>] ("The introduction of structured data requirements for financial statement and prospectus information has had significant benefits for investors, analysts, and other market partici-

from disclosures may be passed on to potential buyers, via a price reduction by competing aggregators.⁷⁰ The cost reduction achieved by automatic extraction of machine-readable disclosures may also allow data aggregators to analyze more disclosures, and collect more information on each disclosure, thereby improving the decision-making by potential buyers.⁷¹ Furthermore, the cost-saving achieved by machine-readable disclosures can reduce the barriers to entry into the data aggregation market, and thereby enable startup aggregators to enter the market more easily.⁷² Increased competition by startup aggregators may contribute to the reduction of prices charged by aggregators to potential buyers.⁷³

b. The Time-Saving Value of Machine-Readable Disclosures

When the data in disclosures is not machine-readable, its collection requires spending considerable time to manually extract the data.⁷⁴ In contrast, machine-readable data allows aggregators to quickly collect the data via an automatic process.⁷⁵ By saving time on manual extraction, automatic extrac-

pants, making it . . . less costly to extract . . . the information in SEC filings.”); Russell G. Golden, Chairman, Fin. Acct. Standards Bd., Remarks at XBRL US Investor Forum 2019: Driving Actionable Analytics (Nov. 4, 2019), <https://www.fasb.org/page/pageContent?pageId=/reference-library/presentationsandspeeches/remarks-of-fasb-chairman-russell-g-golden-xbrl-us-investor.html> [<https://perma.cc/P4AM-LZCN>] (“There are significant advantages to using XBRL data [which is machine-readable] in academic research.”).

⁷⁰ Letter on Investment Company Reporting Modernization, *supra* note 69, at 2 (noting that the cost-savings achieved by automatic extraction of structured data “would be passed on to investors and other users of the data as competition in the data analytics market increases”).

⁷¹ Letter on Fund Retail Investor Experience and Disclosure, *supra* note 58, at 2. When data is machine-readable, analysts are able to gather data in less time and with fewer resources than they could with non-standardized data. *See id.* This results in more data collection overall with respect to the number of companies that analysts can research, as well as the amount of information that they can compile on an individual company. *Id.* As such, the resulting analysis and decisions that derive from that analysis are likely to be of higher quality. *Id.*

⁷² Letter on Investment Company Reporting Modernization, *supra* note 69, at 2 (noting that when data is structured “barriers to entry will decline, enabling startup companies to more easily enter the commercial data business which today (without structured data) must rely on labor-intensive databasing and validation of content”).

⁷³ *Id.* (explaining that increased competition from reducing barriers to entry to data aggregators “can be expected to drive down prices to all investors”).

⁷⁴ Interactive Data to Improve Financial Reporting, Securities Act Release No. 9,002, Exchange Act Release No. 59,324, Trust Indenture Act Release No. 2,461, Investment Company Act Release No. 28,609, 74 Fed. Reg. 6,776, 6,803 (Feb. 10, 2009) (“Investors seeking broad financial coverage of companies must . . . spend considerable time manually collecting the data . . .”).

⁷⁵ Letter on Fund Retail Investor Experience and Disclosure, *supra* note 58, at 2 (noting that machine-readable financial disclosures “can be processed automatically, which . . . makes data available faster”); Mullins, *supra* note 59 (“Data aggregation can enable analysts to access and examine large amounts of data in a reasonable time frame.”); Interactive Data to Improve Financial Reporting, 74 Fed. Reg. at 6,777 (“[R]equiring issuers to file their financial statement information using [the ma-

tion of data from disclosures can increase the timeliness of the data that aggregators present to potential buyers.⁷⁶

c. The Error Reduction Value of Machine-Readable Disclosures

Automatic aggregation of data from disclosures reduces the risk of manual extraction errors, thereby increasing the accuracy of summary information aggregators present to potential buyers.⁷⁷ The accurate presentation of summary information to potential buyers may allow them to compare different products and services in a relevant industry or sector effectively.⁷⁸ The precise aggregation and presentation of data may also reveal industry trends, breakdowns, outliers and historic valuations, which may assist potential buyers in making a superior buying decision.⁷⁹ The accurate presentation of data may also assist data scientists to “apply predictive analytics, artificial intelligence (AI) or machine learning algorithms to the collected data for new insights.”⁸⁰ These insights can ultimately assist potential buyers in making better-informed purchasing decisions.

chine-readable] interactive data format will enable investors, analysts, and the Commission staff to capture and analyze that information more quickly . . .”).

⁷⁶ Letter on Investment Company Reporting Modernization, *supra* note 69, at 2 (stating that a benefit of structured data is “increased timeliness”); Michael S. Piwowar, Comm’r, U.S. Sec. & Exch. Comm’n, Remarks at the 2018 RegTech Data Summit—Old Fields, New Corn: Innovation in Technology and Law (Mar. 7, 2018), <https://www.sec.gov/news/speech/piwowar-old-fields-new-corn-innovation-technology-law> [<https://perma.cc/J5M9-THLW>] (remarking that machine-readable data improve the “timeliness of information processing for . . . consumers of information”).

⁷⁷ Mullins, *supra* note 59 (“[M]annual aggregation risks accidental omission of crucial data sources and patterns.”); Letter on Fund Retail Investor Experience and Disclosure, *supra* note 58, at 2 (“[D]ata that is not available in structured format, must be scraped by data providers and then reviewed, which increases the possibility of introducing errors into the final dataset.”).

⁷⁸ *Cf.* Kara M. Stein, Comm’r, U.S. Sec. & Exch. Comm’n, From the Data Rush to the Data Wars: A Data Revolution in Financial Markets (Sept. 27, 2018), <https://www.sec.gov/news/speech/speech-stein-092718> [<https://perma.cc/H2NK-WB2J>] (“[I]nvestors, with the help of data aggregation platforms and tools, can ultimately use this information to compare investments . . .”); Mark J. Flannery, Chief Economist & Dir., U.S. Sec. & Exch. Comm’n, The Commission’s Production and Use of Structured Data (Sept. 30, 2014), <https://www.sec.gov/news/speech/2014-spch093014mjf> [<https://perma.cc/V9WK-GACP>] (“Without reliable structured data for small companies, data aggregators . . . and investors will be less able to compile the relevant information for the purpose of cross-company analyses.”).

⁷⁹ APIS, *Advantages of Using Aggregated Data APIs in Your Business*, BBVA API MKT. (Jan. 17, 2020), <https://www.bbvaapimarket.com/en/api-world/advantages-using-aggregated-data-apis-your-business/> [<https://perma.cc/P8X7-3V84>] (“When thousands, hundreds of thousands or millions of rows are aggregated in a dataset, you begin to see trends, fragmentation and historic valuations that can be the key to a better understanding of the market for your business . . .” (emphasis omitted)).

⁸⁰ Mullins, *supra* note 59.

d. The Value of Disclosures to Law Enforcement and Legal Scholars

Disclosures in a machine-readable format may not only assist data aggregators and their customers, but also can generate value for disclosure to law enforcement agencies.⁸¹ These agencies can use machine-readable big data, extracted automatically from various long disclosures, to detect “anomalous patterns” that may indicate fraud by disclosers.⁸² Such detection can be done, for example, by identifying discrepancies between “narrative and numeric” aggregated disclosure data.⁸³ It can also be accomplished by comparing a single company’s aggregated disclosure data at different time periods.⁸⁴ Detecting disclosure anomalies may ultimately lead to important inquiries and enforcement actions against fraudulent firms.⁸⁵

Machine-readable disclosures have another important social advantage; they create value for empirical legal scholars.⁸⁶ When disclosures are not ma-

⁸¹ See, e.g., Michael S. Piwowar, Acting Chairman, Sec. & Exch. Comm’n, Opening Statement on Inline XBRL Filing of Tagged Data (Mar. 1, 2017), <https://www.sec.gov/news/statement/opening-statement-on-inline-xbrl-filing-of-tagged-data.html> [<https://perma.cc/UNJ6-66JE>] (“[The SEC] regularly uses XBRL data [which is machine-readable] for analysis in support of . . . enforcement activities.”).

⁸² See Rick A. Fleming, Inv. Advoc., U.S. Sec. & Exch. Comm’n, Improving Disclosure with Smart Data (Oct. 24, 2016), <https://www.sec.gov/news/speech/improving-disclosure-with-smart-data.html> [<https://perma.cc/R3U3-7V3E>] (referring to how the SEC’s Division of Enforcement uses machine-readable data “to detect anomalous patterns in financial statements that may warrant further inquiry” (citing Andrew Ceresney, Dir., Div. of Enf’t, U.S. Sec. & Exch. Comm’n, Directors Forum 2016 Keynote Address (Jan. 25, 2016), <https://www.sec.gov/news/speech/directors-forum-keynote-ceresney.html> [<https://perma.cc/GD74-FXE9>])); Scott W. Bauguess, Acting Dir. & Acting Chief Economist, Div. of Econ. & Risk Analysis, U.S. Sec. & Exch. Comm’n, The Role of Big Data, Machine Learning, and AI in Assessing Risks: a Regulatory Perspective (June 21, 2017), <https://www.sec.gov/news/speech/bauguess-big-data-ai> [<https://perma.cc/EW22-KTHS>] (“Machine learning algorithms may help [the SEC] examiners by pointing them in the right direction in their identification of possible fraud or misconduct . . .”); S.P. Kothari, Chief Economist & Dir., Div. of Econ. & Risk Analysis, U.S. Sec. & Exch. Comm’n, Policy Challenges and Research Opportunities in the Era of Big Data (July 13, 2019), <https://www.sec.gov/news/speech/policy-challenges-research-opportunities-era-big-data> [<https://perma.cc/PZ6W-R8EV>] (“Structuring disclosures so that they are machine readable . . . reduce[s] the ability of filers to hide fraud.”).

⁸³ See Fleming, *supra* note 82 (“[T]he [SEC] has a text analytics initiative in which staff can identify . . . discrepancies between narrative and numeric disclosures . . .”).

⁸⁴ Cf. Elizabeth P. Gray & Catherine E. Fata, *The SEC’s Use of Big Data in the Next Decade: Where Are We Now?*, DIMENSIONS, 2020, no. 1, at 2, 5, <https://www.willkie.com/-/media/files/publications/2020/02/janfeb2020.pdf> [<https://perma.cc/3DVR-UZPP>] (stating that the SEC can “easily aggregate [machine-readable] filing data to compare one company’s disclosures across timeframes”).

⁸⁵ Cf. Fleming, *supra* note 82 (citing Ceresney, *supra* note 82) (describing the SEC Division of Enforcement’s detection of irregular patterns through data aggregation).

⁸⁶ See Stein, *supra* note 78 (“Academia . . . can use this type of [machine-readable] data to, for example, conduct research more quickly and to answer difficult and complex questions about systematic risk at the aggregated level with much less effort.”); Lee, *supra* note 69 (“Structured data in our filings also enables robust academic study.” (first citing Nerissa C. Brown, Adrienna A. Huffman & Shira Cohen, Accounting Reporting Complexity and Non-GAAP Earnings Disclosure 8 n.5 (Dec.

chine-readable, academics' empirical analysis of disclosures may be limited to data from smaller, non-inclusive samples of disclosures.⁸⁷ A potential explanation for academics' reliance on limited data is that manually collecting and processing data from non-machine readable disclosures is a financial burden.⁸⁸ Conversely, when disclosures are machine-readable, empirical legal scholars are usually able to extract data automatically, via computer software, from lengthy and numerous disclosures, thereby saving significant labor-intensive manual extraction costs.⁸⁹ Importantly, the empirical legal studies of disclosures that academics conduct can ultimately benefit policy-makers. For example, when drafting the tentative draft of Restatement of the Law Consumer Contracts, the American Law Institute (ALI), which greatly influences courts and legislators,⁹⁰ relied, *inter alia*, on empirical academic research.⁹¹

II. THE EMPIRICAL TEST

In light of the various social benefits of machine-readable disclosures to data aggregators, their customers, enforcement agencies, and academics, this Part of the Article empirically examines whether disclosures are, in fact, machine-readable. It focuses, as a case study, on testing the machine readability of precontractual disclosures provided by franchisors to prospective franchisees, in the important fast food restaurant franchise industry. Section A of this Part provides an overview of the quick service restaurant franchise industry and its various legal obligations.⁹² Section B describes the sample that forms the basis

2021) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3224798# [<https://perma.cc/EU4N-KG2S>]; and then citing Golden, *supra* note 69)).

⁸⁷ See, e.g., Christopher R. Drahozal & Quentin R. Wittrock, *Is There a Flight from Arbitration?*, 37 HOFSTRA L. REV. 71, 90 (2008) (testing a non-exhaustive sample of seventy-five franchise disclosure documents); Christopher R. Drahozal & Keith N. Hylton, *The Economics of Litigation and Arbitration: An Application to Franchise Contracts*, 32 J. LEGAL STUD. 549, 562 (2003) (testing a similarly limited sample of seventy-five franchise disclosure documents); Benoliel & Zheng, *supra* note 10, at 246 (empirically examining a non-exhaustive sample of 523 U.S. franchise disclosure documents).

⁸⁸ For the costs of analyzing non-machine-readable data, see *supra* notes 64–68 and accompanying text.

⁸⁹ For a discussion of the cost-saving features of machine-readable data, see *supra* note 69 and accompanying text.

⁹⁰ See *About ALI*, AM. L. INST., <https://www.ali.org/about-ali/> [<https://perma.cc/G5GG-JZ6C>] (“ALI drafts, discusses, revises, and publishes Restatements of the Law, Model Codes, and Principles of Law that are enormously influential in the courts and legislatures . . .”).

⁹¹ See, e.g., RESTATEMENT OF CONSUMER CONTRACTS § 5 reps’ notes (AM. LAW. INST., Tentative Draft 2019) (demonstrating a change in courts’ interpretation of unconscionability by referencing an empirical study by Professors DiMatteo and Rich, showing that “courts tend to focus on substantive, rather than procedural, unconscionability” (citing Larry A. DiMatteo & Bruce Louis Rich, *A Consent Theory of Unconscionability: An Empirical Study of Law in Action*, 33 FLA. ST. U. L. REV. 1067 (2006)).

⁹² See *infra* notes 96–110 and accompanying text.

of the study.⁹³ Section C explains the methodology of the study.⁹⁴ Finally, Section D provides an overview of the results of the study.⁹⁵

A. The Quick Service Restaurant Franchise Industry

Franchising plays a vital role in the U.S. economy. It incorporates about 780,000 establishments.⁹⁶ These establishments, in turn, provide almost 8 million jobs.⁹⁷ Furthermore, they annually produce goods and services worth about \$780 billion, and were predicted to contribute approximately \$477 billion to the national GDP in 2021.⁹⁸ Within the franchise business model, the quick service restaurant industry is the largest.⁹⁹ The industry has approximately 190,000 franchise establishments,¹⁰⁰ employs approximately 3.78 million employees,¹⁰¹ and has a yearly nominal output of approximately \$265 billion.¹⁰²

The quick service restaurant franchise industry is governed by a federal disclosure law, known as the Franchise Rule.¹⁰³ The Rule, enacted by the Federal Trade Commission,¹⁰⁴ requires each franchisor to provide potential franchisees with a Franchise Disclosure Document (FDD).¹⁰⁵ Each FDD must equally contain twenty-three prescribed informational items about the franchise.¹⁰⁶ For example, item number five mandates that the franchisor disclose all the fees that a franchisee will pay for “services or goods received from the franchisor before the franchisee’s business opens,” such as training and software fees.¹⁰⁷ As another example, item number seven requires the franchisor to

⁹³ See *infra* notes 111–117 and accompanying text.

⁹⁴ See *infra* notes 118–154 and accompanying text.

⁹⁵ See *infra* notes 155–179 and accompanying text.

⁹⁶ FRANDATA & INT’L FRANCHISE ASS’N, INTERNATIONAL FRANCHISE ASSOCIATION 2021: ECONOMIC OUTLOOK FOR FRANCHISING 6 (2021), https://www.franchise.org/sites/default/files/2021-02/Economic%20Outlook%202021_web2.pdf [<https://perma.cc/22NL-XVT2>].

⁹⁷ *Id.* at 7.

⁹⁸ *Id.* at 1, 8. In 2020, with the COVID-19 pandemic restricting the economy, overall economic output from franchises dropped sharply to an estimated \$670 billion, compared to \$787.5 billion in 2019. *Id.* at 8.

⁹⁹ See FRANDATA & INT’L FRANCHISE ASS’N, *supra* note 96, at 8, 13.

¹⁰⁰ *Id.* at 6.

¹⁰¹ *Id.* at 9, 13.

¹⁰² *Id.* at 10.

¹⁰³ See 16 C.F.R. pts. 436–37 (2022).

¹⁰⁴ The Rule went into effect in 1979 and was amended in 2007. See FED. TRADE COMM’N, FRANCHISE RULE 16 C.F.R. PART 436 COMPLIANCE GUIDE, at i (2008), <https://www.ftc.gov/system/files/documents/plain-language/bus70-franchise-rule-compliance-guide.pdf> [<https://perma.cc/U4UZ-N3D7>].

¹⁰⁵ 16 C.F.R. § 436.2(a) (2022).

¹⁰⁶ See *id.* § 436.5(a)–(w).

¹⁰⁷ *Id.* § 436.5(e); see *Buy a Franchise: What a Franchise Prospect Needs to Know About an FDD*, DRUMMLAW, <https://drummlaw.com/what-is-an-fdd-franchisee-edition/> [<https://perma.cc/7TF6->

disclose the expenses that franchisees should expect to spend during the establishment of their franchise unit (for example, real property, equipment, and inventory expenses).¹⁰⁸ Importantly, pursuant to item number twenty-one, Financial Statements—the focus of this empirical study—a franchisor must disclose their recent balance sheets, *inter alia*.¹⁰⁹ These sheets must be “prepared according to United States generally accepted accounting principles” and “be audited by an independent certified public accountant.”¹¹⁰

B. Sample

The sample of this study includes one hundred FDDs of leading U.S. quick service restaurant franchisors. The FDDs were obtained in two stages. First, I tried to locate as many significant quick service restaurant chains operating in the United States as possible. For that purpose, I mainly used the *Entrepreneur Magazine 2021* dataset.¹¹¹ This dataset includes a list of 1,116 U.S. franchisors from various industries, including the quick service restaurant industry.¹¹² *Entrepreneur* is widely used as a source of data for empirical legal research on franchising.¹¹³ Using this dataset, I located 216 quick service restaurant franchisors.¹¹⁴

Q5PY] (“[The fees disclosed in Item 5] generally include initial franchise fees, any territory fees, training fees, software fees and fees for inventory purchases.”).

¹⁰⁸ 16 C.F.R. § 436.5(g)(1)(i)(C), (D), (E).

¹⁰⁹ *Id.* § 436.5(u)(1)(i). For more details about the balance sheet disclosure and its importance see *supra* Parts II.C.1 and II.C.2, respectively.

¹¹⁰ 16 C.F.R. § 436.5(u)(1).

¹¹¹ See generally *Franchise 500*, ENTREPRENEUR MAG., Jan.–Feb. 2021, at 154 (providing the Franchise 500 Report). In order to complement *Entrepreneur*’s 2021 dataset, I used other sources as well, including the QSR 50 and the Top 500 Chain Restaurant Report. See *Ranking the Top 50 Fast-Food Restaurants in America*, QSR MAG. (2020), <https://www.qsrmagazine.com/content/qs50-2020-top-50-chart> [<https://perma.cc/MD8F-9DQW>] (providing the QSR 50 top performing fast-food quick service restaurants); *Top 500 Chains*, REST. BUS., <https://www.restaurantbusinessonline.com/top-500-2020> [<https://perma.cc/XT3U-DP8P>] (providing the Top 500 Chain Restaurant Report).

¹¹² *Understanding the Ranking*, ENTREPRENEUR MAG., Jan.–Feb. 2021, at 148, 148. Notably, the dataset also ranks the leading five hundred franchises based on various factors such as costs, fees, size, and growth. See *id.*

¹¹³ See, e.g., Drahozal & Wittrock, *supra* note 87, at 90 (using *Entrepreneur*’s rankings for a franchise sample); Drahozal & Hylton, *supra* note 87, at 562 (same); Christopher R. Drahozal & Erin O’Hara O’Connor, *Unbundling Procedure: Carve-Outs from Arbitration Clauses*, 66 FLA. L. REV. 1945, 1979–80 (2014) (same); Christopher R. Drahozal, “Unfair” Arbitration Clauses, 2001 U. ILL. L. REV. 695, 723–24 (same); Bruce H. Kobayashi & Larry E. Ribstein, *Contract and Jurisdictional Freedom*, in THE FALL AND RISE OF FREEDOM OF CONTRACT 325, 344 (F.H. Buckley ed., 1999) (same); Peter B. Rutledge & Christopher R. Drahozal, “Sticky” Arbitration Clauses? *The Use of Arbitration Clauses After Concepcion and Amex*, 67 VAND. L. REV. 955, 988 (2014) (same).

¹¹⁴ *Franchise 500*, *supra* note 111, at 168–78.

Second, within the quick service restaurant chains that I located, I identified the one hundred largest chains¹¹⁵ with FDDs that are publicly available in a major government database, namely the Wisconsin Department of Financial Institutions.¹¹⁶ These chains served as the final sample for this empirical study. Appendix A includes all the sample chains.¹¹⁷

C. Methodology

Within the FDDs in the sample, I tested the machine readability of the franchisors' balance sheets. In order to provide context, I will first briefly explain what the franchisor balance sheet is. I will then explain why the franchisor balance sheet served as the basis of this empirical test. I will finally explain the methodology used to test the machine readability of the balance sheet.

1. The Franchisor Balance Sheet—An Overview

Balance sheets are among the key financial statements that one may use to evaluate a franchise.¹¹⁸ The sheet itself includes data from a specific point in time that concerns three major variables: the franchisor's *assets*, *liabilities*, and *shareholders' equity*.¹¹⁹

The franchisor's *assets* are defined as things that have value, such as buildings and equipment,¹²⁰ and are owned by the franchisor.¹²¹ The assets can

¹¹⁵ See *id.* The size of a chain was based, in this study, on the number of franchised restaurants in the chain. *Id.*

¹¹⁶ *Franchise Search*, WIS. DEP'T OF FIN. INSTS., <https://www.wdfi.org/apps/FranchiseSearch/MainSearch.aspx> [<https://perma.cc/BYR5-AM7Q>]. Notably, Wisconsin is one of a few states that make the FDDs of franchisors available online. See Mike Drumm, *How to Find Franchise Disclosure Documents (FDDs) Online for Free*, DRUMMLAW (May 27, 2019), <https://drummlaw.com/blog/how-to-find-franchise-disclosure-documents/> [<https://perma.cc/7ZYA-L49R>] (“There are four states [including Wisconsin] that publish FDDs online.”).

¹¹⁷ The sample FDDs that I examined were issued by the franchisors in 2021. For the quick service chains that were examined, see Appendix A, located on pages 1946–1947.

¹¹⁸ Jason Fernando, *Balance Sheet*, INVESTOPEDIA, <https://www.investopedia.com/terms/b/balancesheet.asp> [<https://perma.cc/U339-35JC>] (Jan. 1, 2021) (“The balance sheet is one of the three core financial statements that are used to evaluate a business.”).

¹¹⁹ JOEL R. BUCKBERG & ROBERT M. EINHORN, AMERICAN BAR ASSOCIATION 36TH ANNUAL FORUM ON FRANCHISING: THE USE & ANALYSIS OF FINANCIAL STATEMENTS 14 (2013), https://www.americanbar.org/content/dam/aba/events/franchising/2013/w10_2013.pdf [<https://perma.cc/KK5N-46GK>].

¹²⁰ See Adam Barone, *Asset*, INVESTOPEDIA, <https://www.investopedia.com/terms/a/asset.asp> [<https://perma.cc/ZGV5-5L22>] (Jan. 27, 2022) (“An asset can be thought of as something that, in the future, can generate cash flow, reduce expenses, or improve sales . . .”).

¹²¹ Harold Averkamp, *Accounting Basics (Explanation): Balance Sheet—Assets*, ACCOUNTING COACH, <https://www.accountingcoach.com/accounting-basics/explanation/3> [<https://perma.cc/23PP-UNNJ>] (“Assets are things that a company owns and are sometimes referred to as the resources of the company.”).

be categorized into current and noncurrent assets.¹²² Current assets (for example, short-term investments) can be readily transformed into cash in one year or less,¹²³ but noncurrent assets (for example, intellectual property) cannot.¹²⁴

The franchisor's *liabilities* are debts owed by the franchisor to others.¹²⁵ They can be divided, similarly to assets, into current liabilities and long-term liabilities (also known as non-current liabilities).¹²⁶ Current liabilities are the company's liabilities (for example, short term loans) owed within one year of the date on the balance sheet.¹²⁷ Long-term liabilities, however, "are due after a period of at least one year from the date of the balance sheet."¹²⁸

Shareholders' equity "is the remaining amount of assets available to [the franchisors'] shareholders after all liabilities have been paid" off.¹²⁹ If the shareholders' equity is negative, then the liabilities of the company are greater than its assets.¹³⁰ If the equity is positive, however, the company's assets are sufficient to account for its liabilities.¹³¹

In addition to the three major components of the franchisor's balance sheet—assets, liabilities, and shareholder's equity—the sheet often also includes notes to the financial statement.¹³² These notes provide supplementary information that may be needed to clarify, in certain circumstances, the finan-

¹²² Will Kenton, *Noncurrent Assets*, INVESTOPEDIA, <https://www.investopedia.com/terms/n/noncurrent-assets.asp> [<https://perma.cc/67F8-FUPK>] (Sept. 5, 2021). ("A company's assets are divided into two categories: noncurrent and current assets, which appear on a company's balance sheet.")

¹²³ Joshua Kennon, *Understanding Current Assets on the Balance Sheet*, THE BALANCE, <https://www.thebalance.com/current-assets-on-the-balance-sheet-357272> [<https://perma.cc/P9VP-N4GK>] (Oct. 21, 2021).

¹²⁴ Kenton, *supra* note 122 ("Noncurrent assets are a company's long-term investments that are not easily converted to cash or are not expected to become cash within an accounting year.")

¹²⁵ Fernando, *supra* note 118 ("A liability is any money that a company owes to outside parties . . .").

¹²⁶ Harold Averkamp, *Balance Sheet (Explanation): Liabilities, Current Liabilities*, ACCOUNTING COACH, <https://www.accountingcoach.com/balance-sheet-new/explanation/5> [<https://perma.cc/A25N-FV6T>] ("The balance sheet reports two major categories or classifications of liabilities: [c]urrent liabilities [and] [l]ong-term liabilities.")

¹²⁷ *Id.* ("Current liabilities . . . [are] due within one year . . .") (emphasis omitted).

¹²⁸ The Investopedia Team, *Reading the Balance Sheet*, INVESTOPEDIA, <https://www.investopedia.com/articles/04/031004.asp#:~:text=A%20company's%20balance%20sheet%2C%20also,of%20any%20company's%20financial%20statements>. [<https://perma.cc/Z3V3-X39Y>] (Feb. 20, 2021).

¹²⁹ Adam Hayes, *Stockholders' Equity*, INVESTOPEDIA, <https://www.investopedia.com/terms/s/stockholdersequity.asp> [<https://perma.cc/7FA8-Q6PF>] (Apr. 11, 2021).

¹³⁰ Adam Hayes, *Shareholder Equity (SE)*, INVESTOPEDIA, <https://www.investopedia.com/terms/s/shareholdersequity.asp> [<https://perma.cc/FKN5-SUS2>] (Mar. 2, 2022).

¹³¹ *Id.* (explaining the significance of positive and negative shareholders' equity).

¹³² Harold Averkamp, *Balance Sheet (Explanation): How the Balance Sheet and Income Statement Are Connected, Notes to the Financial Statements, Making Sure Your Company's Balance Sheet Is Accurate*, ACCOUNTINGCOACH, <https://www.accountingcoach.com/balance-sheet-new/explanation/7> [<https://perma.cc/E5KJ-JKQR>] ("The notes to the financial statements are an integral (essential) part of the balance sheet." (emphasis omitted)).

cial status of the franchisor.¹³³ To illustrate, a note to the financial statement may disclose occurrences that postdated the balance sheet, but preceded the franchise disclosure.¹³⁴

2. Why Test Balance Sheet Disclosure?

Out of the twenty-three items contained in each franchise disclosure, this study focuses on the machine readability of the balance sheet for three main reasons. First, the process of examining whether all the disclosure items across all the sample FDDs are machine-readable is too cumbersome. Each FDD, including its twenty-three disclosure items, is lengthy and includes hundreds of pages.¹³⁵

Second, the franchisor balance sheet is very important for potential franchisees. This is because the balance sheet itself serves as a useful outline of the franchisor's financial health.¹³⁶ In particular, the balance sheet reflects the franchisor's solvency.¹³⁷ To illustrate, if the franchisor's shareholders' equity is negative, it could raise a red flag for potential franchisees who consider investing in the franchise.¹³⁸ This is because the negative equity indicates that the franchisor's liabilities are greater than its assets.¹³⁹ If the negative equity ultimately leads to the franchisor's bankruptcy, the franchisor might be unable to meet its contractual obligations to its franchisees.¹⁴⁰ These obligations may

¹³³ *Notes to Financial Statements*, BDC, <https://www.bdc.ca/en/articles-tools/entrepreneur-toolkit/templates-business-guides/glossary/notes-to-financial-statements> [<https://perma.cc/XQ56-RGYC>] (“Notes to the financial statements disclose the detailed assumptions made by accountants when preparing a company’s . . . balance sheet . . .”).

¹³⁴ *Financial Statement Notes*, CORP. FIN. INST., <https://corporatefinanceinstitute.com/resources/knowledge/accounting/financial-statement-notes/> [<https://perma.cc/2JFR-LY33>] (“Information on any subsequent events can be found also in the financial statement notes section. Subsequent events refer to events that occur after the balance sheet date but before the release of the financial statements.”).

¹³⁵ See *supra* note 65 and accompanying text for a discussion of the various lengths of FDDs.

¹³⁶ Nick Powills, *Understanding FDD Item 21: Audited Financial Statements*, 1851 FRANCHISE (June 23, 2017), <https://1851franchise.com/understanding-fdd-item-21-audited-financial-statements-2703110#stories> [<https://perma.cc/688X-XM3G>]; BUCKBERG & EINHORN, *supra* note 119, at 19 (“The balance sheet is a snapshot of the financial health of the enterprise as of its date.”).

¹³⁷ *Item 21: Financial Statements*, THE EDUCATED FRANCHISEE FDD EXCH., <https://fdexchange.com/fdd-information/franchise-disclosure-document-checklist/item-21-financial-statements/> [<https://perma.cc/26AF-RPDY>] (explaining that the franchisors' balance sheets “give the franchise prospect a clear picture of the solvency of the franchisor”).

¹³⁸ See Ken Clark, *What Does Negative Shareholders' Equity Mean?*, INVESTOPEDIA, <https://www.investopedia.com/ask/answers/08/negative-shareholder-equity.asp> [<https://perma.cc/FZQ4-XH5L>] (Apr. 6, 2022) (explaining that negative shareholders' equity may suggest to prospective investors that a business is in poor financial health).

¹³⁹ *Id.*

¹⁴⁰ See *Franchisor Insolvency—It May Happen*, MORGAN MAC LAWS., <https://www.morganmac.com.au/franchisor-insolvency-it-may-happen/> [<https://perma.cc/EL88-GR6X>] (describing consequences of franchisor insolvency, including the potential that “the liquidator of the insolvent franchisor may

include, for example, providing franchisees with aid, advice, and training, as well as further developing the franchise chain.¹⁴¹

Moreover, the federal Franchise Rule requires franchisors to disclose their balance sheets for the prior two fiscal years.¹⁴² This is crucial because prior balance sheets assist prospective franchisees in evaluating shifts in the franchisor's financial well-being.¹⁴³ For example, a positive trend in the value of the franchisor's assets and equity and a negative trend with respect to the franchisor's liabilities may suggest improvement in the franchisor's financial health.¹⁴⁴ In contrast, a negative trend in the value of the franchisor's assets and equity and a positive trend in the value of liabilities on the franchisor's balance sheet may indicate a deterioration in financial health.¹⁴⁵

Finally, this study focuses on balance sheet disclosure for a third reason: balance sheet data are important to franchise enforcement entities. In many states, official examiners will analyze a prospective franchisor's balance sheet before allowing the franchisor to sell franchises.¹⁴⁶ The examiners might reject franchisors with unsatisfactory financial statements.¹⁴⁷ Alternatively, the state might only approve a franchisor contingent on the franchisor's satisfaction of certain financial precautionary measures.¹⁴⁸ These safeguards may include, for example, placing the initial franchise fees that franchisees usually pay in escrow until the franchisor has fulfilled its initial responsibilities to its franchisees (for example, site selection assistance or pre-opening training).¹⁴⁹

not continue with the franchise agreements"); *How Bankruptcy Affects Franchise Agreements*, KANE RUSSELL COLEMAN LOGAN (Oct. 8, 2014), <https://insolvencyinsights.com/2014/10/09/how-bankruptcy-affects-franchise-agreements/> [<https://perma.cc/V6L8-8WR5>] ("The bankruptcy filing of a franchisor . . . can greatly increase the uncertainty of how and when the provisions of a franchise agreement will be enforced.").

¹⁴¹ John Pratt, *What Are a Franchisor's Obligations to Its Franchise Network?*, WHAT FRANCHISE, <https://www.what-franchise.com/questions/what-are-a-franchisors-obligations-to-its-franchise-network> [<https://perma.cc/7ZCR-APYJ>].

¹⁴² See 16 C.F.R. § 436.5(u)(i) (2022).

¹⁴³ BUCKBERG & EINHORN, *supra* note 119, at 19 ("The customary presentation compares the balance sheet values to the values from the prior year so the reader can assess instantly the directional changes in the financial health of the enterprise.").

¹⁴⁴ *Id.*

¹⁴⁵ *Id.*

¹⁴⁶ Powills, *supra* note 136 (describing the review processes some states take to approve or deny a franchisor attempting to market franchises in those states). These states supplement the federal government's requirements, providing prospective franchisees with additional safeguards. BUCKBERG & EINHORN, *supra* note 119, at 34.

¹⁴⁷ Powills, *supra* note 136 ("A franchisor with weak financial statements may be rejected by certain states. . .").

¹⁴⁸ *Id.*

¹⁴⁹ *Id.*

3. How Was Machine Readability Tested?

To test the machine readability of balance sheet disclosures, I thoroughly reviewed the data and structure of the sheets included in all the one hundred sample FDDs. While reviewing these sheets, I checked whether their data could be easily and accurately identified and extracted by a computer. For that purpose, I specifically examined four questions, which I developed in light of the typical characteristics of machine-readable data:¹⁵⁰

1. Are the balance sheets provided in a *digital format* that technically enables a computer to easily search and extract data included in the sheets? On one hand, if the balance sheet data is in text format within a PDF file, a computer can easily search and extract the data.¹⁵¹ Conversely, if the text of the balance sheet is in a scanned image format within a PDF file, a computer cannot simply and correctly identify and extract the data. For a computer to extract the data, first, the scanned image needs to be converted to searchable text via OCR method, which is prone to errors (for example, character exchanges).¹⁵²
2. Do all balance sheets have a *uniform structure* that can assist a computer to extract data systematically from identical parts within this structure? In order to answer this question, three sub-questions were examined:
 - (a) Does the structure of the balance sheets uniformly and explicitly include the three major parts that represent the sheet's main components: total assets, total liabilities and total shareholders' equity?¹⁵³
 - (b) Does the structure of the balance sheets consistently and explicitly include the major subparts of the sheet? More specifically, does the balance sheet explicitly include the two major subparts that represent the central components of the assets item: current assets and non-current assets?¹⁵⁴ Additionally, do

¹⁵⁰ For the characteristics of machine-readable disclosures, see *supra* Subsection I.B.1.

¹⁵¹ Sean Broderick, *Three Types of PDFs*, NAT'L LITIG. SUPPORT BLOG (June 12, 2020), <https://nlsblog.org/2020/06/12/three-types-of-pdfs/> [<https://perma.cc/E5LT-E9F2>] (explaining that in text-based PDFs, "[t]he text layer is searchable text"); see *Adobe PDF Extract API*, ADOBE DEV., <https://www.adobe.io/apis/documentcloud/dcsdk/pdf-extract.html> [<https://perma.cc/T2DP-KGWF>] (noting that text from PDF files can be extracted using a "structured" text-based file).

¹⁵² For a discussion of OCR technology and its drawbacks, see *supra* note 55 and accompanying text.

¹⁵³ For the major components of a franchisor balance sheet, see *supra* Subsection II.C.1.

¹⁵⁴ For the major components of the assets item, see *supra* Subsection II.C.1.

the balance sheets include a distinction between current and other, non-current, liabilities?

- (c) Are the supplementary notes of the balance sheets presented in a uniform manner, within the balance sheet structure?
3. Do the balance sheets use a *standardized taxonomy* that can assist computers to extract and process data systematically from uniform data items within the sheets? In this respect, two sample questions were examined:
- (a) Is the taxonomy used in the sheets to represent the shareholders' equity uniform?
- (b) Is the numeric baseline (i.e., thousands, millions, or actual numbers) of the financial data in the balance sheets uniform?
4. Do the balance sheets apply *unique*, non-duplicate, data identifiers that distinguish between different data items? In this respect, two sample questions were examined:
- (a) Is the term "total assets" used only once in each disclosure?
- (b) Is the term "total liabilities" used only once in each disclosure?

D. Results

The study's findings alarmingly indicate that the balance sheet disclosures are non-machine-readable. Here are the detailed results:

1. Most Balance Sheets Are in a Digital Format That Allows for Some Level of Data Searching and Extraction

Fortunately, most of the balance sheets are in a *digital format* that allows a computer—from a narrow technical perspective—to search and extract data included in the sheets easily. Particularly, the vast majority of the balance sheets (91%) are in a searchable and extractable text format within a PDF file. A non-trivial percentage of the balance sheets, however, namely 9%,¹⁵⁵ are scanned images within PDF files. Unfortunately, these images cannot be easily

¹⁵⁵ The following disclosures provide the balance sheet in a scanned image format: Different Rules, LLC dba Jack in the Box, Franchise Disclosure Document 92 (2021); Nektar Franchise, Inc. dba Nektar Juice Bar, Franchise Disclosure Document 64 (2021); Kona Ice, Inc., Franchise Disclosure Document 63 (2021); Pizza Ranch, Inc., Franchise Disclosure Document 148 (2021); Chester's Int'l, LLC, Franchise Disclosure Document 63 (2021); FSC Franchise Co., LLC dba Beef 'O' Brady's, Franchise Disclosure Document 69 (2021); Blaze Pizza, LLC, Franchise Disclosure Document 76 (2021); Wetzel's Pretzels, LLC, Franchise Disclosure Document 89 (2021); Lee's Franchisor, LLC dba Lee's Famous Recipe Chicken 142 (2021).

and accurately read by computers and require an OCR conversion, which is error-prone.¹⁵⁶

2. Balance Sheets Do Not Have a Unified Structure That Would Facilitate Systematic Computerized Data Extraction

For several reasons, the results of this study alarmingly show that the balance sheets do not have a *unified structure* that may facilitate the systematic extraction of data by computers. First, the division of the balance sheet to major components is not uniform. Some balance sheets explicitly include three distinct titles that represent the major building blocks of a balance sheet: “[t]otal assets,” “[t]otal liabilities,” and “[t]otal stockholders’ equity.”¹⁵⁷ Some balance sheets, however, fail to include these distinct titles entirely.¹⁵⁸

Second, the division of the balance sheet to subcomponents is not consistent. Some balance sheets include a clear distinction between the different major subcomponents of the three building blocks of the balance sheet. For example, some balance sheets include a distinction between current and other non-current assets.¹⁵⁹ Likewise, some balance sheets include a distinction between current and other non-current liabilities.¹⁶⁰ Some balance sheets, however, fail to explicitly include such distinctions.¹⁶¹

Third, the notes in the balance sheets, which provide supplementary information to clarify the financial status of the franchisor, are not presented under a uniform structure.¹⁶² For example, some balance sheets include a distinct

¹⁵⁶ For an explanation of why scanned images are not machine-readable, see *supra* notes 52–55 and accompanying text.

¹⁵⁷ See, e.g., Pizza Ranch, Inc., *supra* note 155, at 148.

¹⁵⁸ See, e.g., Freddy’s, LLC dba Freddy’s Frozen Custard & Steamburgers, Franchise Disclosure Document 165–66 (2021) (using titles other than the explicit distinct titles “[t]otal assets,” and “[t]otal stockholders’ equity”); Firehouse of Am., LLC dba Firehouse Subs, Franchise Disclosure Document 101 (2021) (same); Golden Corral Franchising Sys., Inc., Franchise Disclosure Document 284 (2021) (same).

¹⁵⁹ See, e.g., Citadel Panda Express, Inc., Franchise Disclosure Document 204 (2021); A&W Rests., Inc., Financial Disclosure Document 100 (2021); Gosh Enters., Inc. dba Charleys Philly Steaks, Franchise Disclosure Document 152 (2021); A Sub Above, LLC dba Jersey Mike’s, Franchise Disclosure Document 153 (2021); Dippin’ Dots Franchising, LLC, Franchise Disclosure Document 137 (2021).

¹⁶⁰ See, e.g., BurgerFi Int’l, Inc., Franchise Disclosure Document 72 (2021); A&W Rests., Inc., *supra* note 159, at 101; Gosh Enters., Inc. dba Charleys Philly Steaks, *supra* note 159, at 153.

¹⁶¹ See, e.g., Chick-fil-A, Inc., Franchise Disclosure Document 279–80 (2021); Little Caesar Enters., Inc., Franchise Disclosure Document 278 (2021).

¹⁶² For discussion of the purpose and value of notes in balance sheets, see *supra* note 133 and accompanying text.

column with a number that identifies each note.¹⁶³ In these sheets, the number of each note is disclosed alongside the balance sheet item that the note references.¹⁶⁴ To illustrate, Blimpie’s balance sheet includes a distinct column in which a note, numbered nine, is disclosed alongside the balance sheet item entitled “[i]ntangible assets.”¹⁶⁵ However, other balance sheets fail to include this type of distinctive column. Instead, these sheets include a footer with a general statement referring to the notes, such as “[s]ee accompanying notes to consolidated financial statements.”¹⁶⁶

3. Balance Sheets Often Include a Non-Standardized Taxonomy That Impedes Machine Reading

The results of this study disturbingly show that the examined balance sheets often include a non-standardized *taxonomy* that impedes their machine-reading. First, the balance sheets often use non-uniform language to present similar data. To illustrate, the sheets regularly label the shareholders’ equity, a financial item in the sheet, in numerous non-standardized ways such as “[o]wners’ equity,”¹⁶⁷ “[e]quity,”¹⁶⁸ “[m]embers’ equity,”¹⁶⁹ “[m]embers’ capital,”¹⁷⁰ “[s]tockholders’ . . . equity,”¹⁷¹ or “[s]hareholders’ equity.”¹⁷² Second, the numeric baseline of the financial data in the balance sheets is not uniform. Some balance sheets present their financial numbers in the millions,¹⁷³ others in the thousands,¹⁷⁴ and others use actual figures.¹⁷⁵

¹⁶³ See, e.g., Papa Murphy’s Int’l LLC, Franchise Disclosure Document 138–39 (2021); BF Acquisition Holdings, LLC dba Baja Fresh, Franchise Disclosure Document 556–57 (2021); Kahala Franchising, LLC dba Blimpie, Franchise Disclosure Document 597–98 (2021).

¹⁶⁴ See Papa Murphy’s Int’l LLC, *supra* note 163, at 138–39; BF Acquisition Holdings, LLC dba Baja Fresh, *supra* note 163, at 556–57; Kahala Franchising, LLC dba Blimpie, *supra* note 163, at 597–98.

¹⁶⁵ Kahala Franchising, LLC dba Blimpie, *supra* note 163, at 597. As another example, on the balance sheet of Baja Fresh, note number sixteen is alongside an item entitled “[l]ong-term debt.” BF Acquisition Holdings, LLC dba Baja Fresh, *supra* note 163, at 556.

¹⁶⁶ Burger King Corp., Franchise Disclosure Document 718 (2021); see also Glob. Orange Dev., LLC dba Biggby Coffee, Franchise Disclosure Document 289 (2021) (including the general statement “[s]ee notes to consolidated financial statements”).

¹⁶⁷ See, e.g., Rosati’s Franchising, Inc., Franchise Disclosure Document 89 (2021).

¹⁶⁸ See, e.g., Kona Ice, Inc., *supra* note 155, at 64; Captain D’s, LLC, Franchise Disclosure Document 61 (2021).

¹⁶⁹ See, e.g., KFC US, LLC, Franchise Disclosure Document 143 (2021).

¹⁷⁰ See, e.g., Donatos Pizzeria, LLC, Franchise Disclosure Document 92 (2021).

¹⁷¹ See, e.g., Qdoba Rest. Corp., Franchise Disclosure Document 75 (2021).

¹⁷² See, e.g., Popeyes La. Kitchen, Inc., Franchise Disclosure Document 458 (2021).

¹⁷³ See, e.g., Burger King Corp., *supra* note 166, 718.

¹⁷⁴ See, e.g., Jamba Juice Franchisor SPV LLC, Franchise Disclosure Document 103 (2021).

¹⁷⁵ See, e.g., Pizza Ranch, Inc., *supra* note 155, at 148.

4. Disclosures Do Not Employ Unique, Non-Duplicative Data Identifiers

Disturbingly, the results of this study indicate that disclosures do not employ unique, non-duplicate data identifiers. First, the important balance sheet term “total assets” is not unique within each disclosure. For example, this term appears twelve times in different segments of Applebee’s disclosure.¹⁷⁶ Similarly, this term appears nine times in various locations of Tim Hortons’ disclosure.¹⁷⁷ In addition, the important balance sheet term “total liabilities” does not serve as a unique data identifier in each disclosure. For example, this term appears ten times in Dunkin’s disclosure.¹⁷⁸ Likewise, the term appears eight times in the disclosure for Marco’s Pizza.¹⁷⁹

III. NORMATIVE DISCUSSION AND IMPLICATIONS

The results of this study indicate that franchise disclosures are non-machine readable. Notably, the legal framework that underlines the disclosures might serve as an explanation for their non-machine-readability. Namely, the Franchise Rule aims to ensure that disclosures are *human*-readable, but does not necessarily ensure they are *machine*-readable.¹⁸⁰ In order to enhance the human readability of disclosures, the Rule requires franchisors to provide FDDs that are written “clearly, legibly, and . . . [in] plain English.”¹⁸¹ Furthermore, it requires that the language of FDDs be “understandable by a person unfamiliar with the franchise business.”¹⁸² Relatedly, the Rule also requires that FDDs utilize “short sentences” and “everyday language.”¹⁸³

Other federal disclosure laws also focus on ensuring that disclosures are *human*-readable. For example, the Employee Retirement Income Security Act of 1974, which Congress enacted to protect the interests of employee benefit plan participants, requires that the disclosure of the terms and benefits of a plan be “written in a manner calculated to be understood by the average plan participant.”¹⁸⁴ Similarly, the Electronic Fund Transfer Act, which establishes the rights of consumers in electronic fund transfer activities, requires infor-

¹⁷⁶ See Applebee’s Franchisor LLC, Franchise Disclosure Document (2021).

¹⁷⁷ See Tim Hortons USA Inc., Franchise Disclosure Document (2021).

¹⁷⁸ See Dunkin’ Donuts Franchising, LLC, Franchise Disclosure Document (2021).

¹⁷⁹ See Marco’s Franchising, LLC dba Marco’s Pizza, Franchise Disclosure Document (2021).

¹⁸⁰ See, e.g., 16 C.F.R. § 436.6(b) (2022) (requiring disclosures that are readable for “prospective franchisees”).

¹⁸¹ *Id.*

¹⁸² *Id.* § 436.1(o) (defining “[p]lain English” as “the organization of information and language usage understandable by a person unfamiliar with the franchise business”).

¹⁸³ *Id.* (“*Plain English* . . . incorporates short sentences; definite, concrete, everyday language . . .”).

¹⁸⁴ Employee Retirement Income Security Act of 1974, 29 U.S.C. § 1022(a).

mation disclosed to consumers to be “clear and readily understandable.”¹⁸⁵ Furthermore, the Magnuson-Moss Warranty Act requires that the disclosed information about the terms of a consumer product warranty be “simple and readily understood.”¹⁸⁶ In addition, the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, which provides individuals a right to be informed of the privacy practices of their health plans, requires information to be disclosed “in plain language.”¹⁸⁷ Finally, the Truth in Lending Act, which aims to promote the informed use of consumer credit, requires information to “be disclosed clearly” to consumers.¹⁸⁸

Although federal disclosure laws often aim to ensure that disclosures are *human*-readable, they normally fail to guarantee that disclosures are *machine*-readable. These laws, therefore, ignore the significant social benefits of machine-readable disclosures.¹⁸⁹ Accordingly, this Article calls on policy-makers to adopt new disclosure rules that facilitate the readability of disclosures by computers. For this purpose, policy-makers should require disclosure documents, in each relevant industry, to meet four major conditions:

1. Disclosures must be in a *digital format* that technically enables a computer to search and extract data included in the disclosures easily. Accordingly, disclosures must not include text in a scanned image format that is prone to OCR conversion errors.¹⁹⁰
2. Disclosures must have a *unified structure* that enables computers to extract data systematically from identical parts within this structure. For example, franchisor balance sheet disclosures should include, *inter alia*, three distinct and explicit major components: total assets, total liabilities, and total shareholders’ equity.
3. Disclosures must use *standardized taxonomies* that can assist computers to extract and process data systematically from various data items. For instance, in the franchising context, the taxonomy

¹⁸⁵ 15 U.S.C. § 1693(b); 12 C.F.R. §§ 205.1(b), 205.4(a)(1) (2022) (requiring, pursuant to the Electronic Fund Transfer Act, that disclosures “shall be clear and readily understandable . . .”).

¹⁸⁶ 15 U.S.C. § 2302(a) (“[A]ny warrantor warranting a consumer product to a consumer by means of a written warranty shall . . . disclose in simple and readily understood language the terms and conditions of such warranty.”).

¹⁸⁷ See 45 C.F.R. 164.520(b)(1) (2022).

¹⁸⁸ See 15 U.S.C. § 1632(a). For additional disclosure laws, requiring information to be disclosed “clearly” or in a “clear” manner, see 12 C.F.R. § 213.3(a) (2022) (requiring that disclosures be made “clearly” pursuant to the Consumer Leasing Act); 12 C.F.R. § 1024.32(a)(1) (requiring, pursuant to the Real Estate Settlement Procedures Act, “clear and conspicuous” disclosures); 12 C.F.R. § 1030.3(a) (requiring that disclosures be made “clearly” under the Truth in Savings Act); The R-Value Rule, 16 C.F.R. § 460.10 (2022) (requiring “clearly and conspicuously” made statements).

¹⁸⁹ For the benefits of machine-readable disclosures, see *supra* Subsection I.B.2.

¹⁹⁰ For a discussion of the drawbacks of OCR conversion, see *supra* note 55 and accompanying text.

used in balance sheet disclosures to represent the shareholders' equity must be uniform. Similarly, the numeric baseline (i.e., thousands, millions, or actual numbers) of the financial data in the balance sheets must be standardized.

4. The data items that need to be disclosed must be presented in a manner that enables computers to easily locate them within the disclosure. For that purpose, the data items must include uniform *unique* titles, symbols, or tags that differentiates them from other data elements in the disclosure. To illustrate in the context of franchising, the balance sheet disclosed by each franchisor must include a unique title (i.e., "balance sheet").

Remarkably, a disclosure regime similar to the machine readability scheme that is proposed in this Article has already been adopted in limited disclosure areas. Namely, in 2009, the SEC adopted rules governing the securities industry that require publicly traded companies to disclose financial statement information in a machine-readable format.¹⁹¹ Pursuant to these rules, public companies are specifically required to provide their financial statements to the SEC in an interactive data format using the eXtensible Business Reporting Language (XBRL).¹⁹² XBRL is an electronic format that requires disclosure documents to include a standard list of "tags."¹⁹³ This list of tags contains descriptive labels and definitions, which provide the contextual information necessary for computers to identify and process the disclosed data.¹⁹⁴ In explaining the use of tags, the SEC provided a hypothetical example whereby a public company disclosing \$1 million in net sales would need to generate machine-readable data that "identify what the 1,000,000 represents, net sales, and the currency in which it is disclosed, dollars."¹⁹⁵ In the hypothetical disclosure, the company would also provide "contextual information," namely, additional tagged information needed for machine-reading.¹⁹⁶ Such information might include the company's identification number, whether the disclosed data de-

¹⁹¹ See Interactive Data to Improve Financial Reporting, Securities Act Release No. 9,002, Exchange Act Release No. 59,324, Trust Indenture Act Release No. 2,461, Investment Company Act Release No. 28,609, 74 Fed. Reg. 6,776, 6,776 (Feb. 10, 2009) (explaining that, pursuant to the adopted rules, companies must provide financial statement data in a format that "could be downloaded directly into spreadsheets, analyzed in a variety of ways using commercial off-the-shelf software, and used within investment models in other software formats").

¹⁹² *Id.*

¹⁹³ *Id.* at 6,778 (describing how tags, or standardized identifiers for data, allow for streamlined computer-based interpretation of the data).

¹⁹⁴ *Id.* (describing tags as "similar to definitions in an ordinary dictionary . . . [that] cover a variety of financial concepts that can be read and understood by software applications").

¹⁹⁵ *Id.* at 6,778 n.51.

¹⁹⁶ *Id.*

rives from an “annual” or “quarterly” report, and a description of the report’s concrete financial period.¹⁹⁷ The SEC explained that the use of tags in such disclosures creates a “consistent structure of identity and context.”¹⁹⁸ The “consistent structure” allows for computers to identify and process disclosures vis-à-vis various software applications, such as databases, financial reporting systems, and spreadsheets.¹⁹⁹

Ultimately, the use of the XBRL machine-readable format that the SEC required for the securities industry allowed data aggregators and their customers to harness the power of big data. This can be illustrated by the following example: Calcbench is a data aggregator that extracts data from more than twelve thousand machine-readable financial statements disclosed only by public companies to the SEC.²⁰⁰ It aggregates, *inter alia*, “[m]ore than 1,000 standard data points” from financial statements, such as a company’s revenues and unremitted foreign earnings.²⁰¹ The data aggregator quickly processes and presents these data points minutes after a public company provides them to the SEC.²⁰² The collected data is provided by the aggregator to its customers in summary format, namely in Excel sheets.²⁰³ The presentation of the data in Excel sheets can allow customers to filter, sort, and visualize relevant aggregated data using graphs.²⁰⁴ Customers can also use various formulas in the Ex-

¹⁹⁷ *Id.*

¹⁹⁸ *Id.* at 6,778.

¹⁹⁹ *Id.* Similarly, in 2009 the SEC adopted rules that mandated that mutual funds disclose “risk/return summary information” in machine-readable form, namely XBRL. See Interactive Data for Mutual Fund Risk/Return Summary, Securities Act Release No. 9,006, Exchange Act Release No. 59,391, Trust Indenture Act Release No. 2,462, Investment Company Act Release No. 28,617, 74 Fed. Reg. 7,748, 7,748 (Feb. 19, 2009).

²⁰⁰ *Data Overview: Calcbench Data Sets*, CALCBENCH, https://www.calcbench.com/home/our_data [<https://perma.cc/ZM32-F5LY>] (stating that Calcbench datasets cover “[o]ver 12,000 listed public companies”); see Rani Hoitash, Udi Hoitash & Landi Morris, *eXtensible Business Reporting Language (XBRL): A Review and Implications for Future Research*, 40 AUDITING: J. PRAC. & THEORY 107, 132 (2021) (“Calcbench is one such data aggregator that extracts data from XBRL and provides varying sets of information.”); *Tools and Services*, XBRL, <https://www.xbrl.org/the-standard/how/tools-and-services/entry/916/> [<https://perma.cc/NV9N-3YSM>] (“Calcbench is the first company of its kind to fully harness the power of the new government mandated data standard XBRL, yielding an unprecedented direct line into the SEC’s corporate financial data repository.”).

²⁰¹ See *Data Overview: Calcbench Data Sets*, *supra* note 200 (describing the filing information that the Calcbench Standardized Numeric Financials data set contains).

²⁰² Calcbench Support, *How Soon Is 10-K/Q Data Available from Calcbench?*, CALCBENCH, <https://knowledge.calcbench.com/hc/en-us/articles/231183507-How-soon-is-10-K-Q-data-available-from-Calcbench-> [<https://perma.cc/DD5F-4D5N>].

²⁰³ *Excel Add-in: Data & Modeling*, CALCBENCH, <https://www.calcbench.com/home/exceladdin> [<https://perma.cc/4DJ9-PTC8>] (“Calcbench clients can now access hundreds of normalized accounting metrics, full financial statements, and raw XBRL tags for 12,000+ publicly traded companies directly in Excel.”).

²⁰⁴ See Tyrone Pernsley, *Sorting and Filtering Data with Excel*, LEARN EXCEL NOW, <https://www.learnexcelnow.com/sorting-and-filtering-data-with-excel/> [<https://perma.cc/3S5H-BSE2>] (ex-

cel sheets to derive meaningful financial conclusions from the aggregated data.²⁰⁵ Calcbench furthermore provides “side-by-side comparisons of standardized financial metrics,” within various sectors at different time periods.²⁰⁶ In doing so, the data aggregator assists investors in making better-informed investment decisions.²⁰⁷

In light of the social benefits of the SEC’s machine-readability regime adopted so far in the securities industry, policy-makers should adopt similar regimes in other various areas of disclosure, including franchising, credit cards, and mortgages.²⁰⁸

CONCLUSION

Through the lens of a case study on balance sheet disclosure in the important quick service restaurant franchise industry, this Article empirically indicates that franchise disclosures are non-machine readable. First, a computer cannot easily locate the disclosed data within the disclosures’ text. Second, the disclosed data does not have a unified structure that may facilitate the systematic extraction of data by computers. Third, the disclosed data often includes a non-standardized taxonomy that impedes its machine-reading.

When disclosures are non-machine-readable, it deprives society of the potential social benefits of machine-readable data. It particularly curtails the cost-saving, time-saving, and error-reduction benefits that machine-readable data generate to data aggregators and their clients. It also hinders the ability of enforcement agencies to detect, based on machine-reading of disclosures, anomalous patterns that may indicate fraud by disclosers. Furthermore, it impedes the

plaining the sorting and featuring functions in Excel); Sara Davidson, *How to Make a Chart or Graph in Excel [With Video Tutorial]*, HUBSPOT, <https://blog.hubspot.com/marketing/how-to-build-excel-graph#> [<https://perma.cc/4FCE-3BAW>] (Mar. 16, 2022) (explaining the graph functions in Excel).

²⁰⁵ See Press Release, Calcbench, Calcbench Launches an Excel Add-in, Giving Users Greater Flexibility & Ease-of-Use (Sept. 10, 2014), <https://www.calcbench.com/Content/resources/Calcbench%20Excel%20Add%20In%20PR.pdf> [<https://perma.cc/9KGA-2SZV>] (“Users can manually enter formulas in Excel . . .”).

²⁰⁶ See *Multi-Company: Standardized Data*, CALCBENCH, https://www.calcbench.com/home/compare_multiple_companies [<https://perma.cc/DRF9-P4MF>].

²⁰⁷ According to Calcbench, its services are “[u]sed by Fortune 500 companies, major money management firms, auditors, and research firms.” *About Us: Calcbench Team*, CALCBENCH (emphasis omitted), <https://www.calcbench.com/home/aboutus> [<https://perma.cc/W3V6-T2TR>].

²⁰⁸ For a similar proposal, see, for example, Richard H. Thaler & Will Tucker, *Smarter Information, Smarter Consumers*, HARV. BUS. REV., Jan.–Feb. 2013, at 45, 51 (“The logical next step is for the pages of fine print that we now call disclosure to be replaced by machine-readable files in standardized formats.”); RICHARD H. THALER & CASS R. SUNSTEIN, *NUDGE: THE FINAL EDITION* 143 (Penguin Books 3d ed. 2021) (2008) (“[C]omplex information should be disclosed and made available in a format that is both standardized and machine readable.”).

ability of empirical legal scholars to examine, via machine-reading, the big data included in the numerous disclosures available.

Policy-makers willing to allow society to harness the power of the big data included in disclosures should assure that these disclosures are machine-readable. They should particularly define statutory machine-readability standards and monitor disclosers' compliance with these standards. Otherwise, disclosures may continue failing to keep up with the big data revolution.

APPENDIX A: FRANCHISES STUDIED (ALPHABETICALLY LISTED BY COLUMN)

A&W Restaurants	Domino's Pizza	Papa John's
Applebee's	Donatos Pizza	Papa Murphy's
Arby's	Dunkin'	Pinkberry
Auntie Anne's	Erbert & Gerbert's Sandwich Shop	Pizza Hut
Baja Fresh	Firehouse Subs	Pizza Ranch
Baskin-Robbins	Freddy's Frozen Custard & Steakburgers	PJ's Coffee of New Orleans
Beef 'O' Brady's	Fuzzy's Taco Shop	Planet Smoothie
Ben & Jerry's	Golden Corral	Popeyes Louisiana Kitchen
Ben's Soft Pretzels	Goodcents Restaurant	Qdoba Mexican Eats
Biggby Coffee	Great Harvest Bread Co.	Rally's Restaurants
Blaze Pizza	Great Steak	Rosati's Pizza
Blimpie	Haagen-Dazs Shop	Schlotzsky's
Bojangles	Hungry Howie's Pizza & Subs	Scooter's Coffee
Bonchon	IHOP	Slim Chicken's
Buffalo Wild Wings	Jack In The Box	Smashburger
Burger King	Jamba	Smoothie King
BurgerFi	Jason's Deli	Sonic
Capriotti's Sandwich Shop	Jersey Mike's	Steak 'n Shake
Captain D's	Jimmy John's Gourmet Sandwiches	Subway
Caribou Coffee	KFC	SweetFrog
Carvel	Kona Ice	Taco Bell
Charleys Philly Steaks	Lee's Famous Recipe Chicken	Taco John's
Checkers Restaurants	Little Caesars	Taco Time
Chester's	Marco's Pizza	Teriyaki Madness
Chick-fil-A	Maui Wowi	Texas Roadhouse
Church's Chicken	McAlister's Deli	The Human Bean

Cinnabon	McDonald's	Tim Hortons
Clean Juice	Mellow Mushroom Pizza	Tropical Smoothie Cafe
Cold Stone Creamery	Moe's Southwest Grill	Wayback Burgers
Cousins Subs	MOOYAH®	Wendy's
Culver's	Nathan's Famous	Wetzel's Pretzels
Dairy Queen	Nekter Juice Bar	Zoup Eatery
Denny's Restaurant	Newk's Eatery	
Dippin' Dots	Panda Express	

