Open Data for Financial Reporting: Costs, Benefits, and Future

By Marc D. Joffe
INTRODUCTION

In 2009, the Securities and Exchange Commission (SEC) began requiring public companies to file mandatory disclosures in eXtensible Business Reporting Language (XBRL). XBRL documents are “machine-readable” files that contain text delimited within electronic tags. The use of these tags — which are specified by an SEC-mandated data dictionary — allows the text to be readily loaded into computer databases, spreadsheets, or other software applications. This facilitates further analysis by regulators and other consumers of regulatory filings. In some cases, the SEC requests filings in eXtensible Markup Language (XML), a more general structured data standard from which XBRL was derived.

Both XBRL and XML transform information from document-based forms into a standardized electronic structure, allowing the SEC to collect and publish its regulatory filings as open data.

This study discusses the history, benefits, drawbacks, and future of mandates for regulated entities to report financial information to regulatory agencies as open data with the goal of informing policy debates about whether these requirements should be extended, curtailed, or even abolished. Most of the discussion will focus on the SEC’s requirement that publicly listed companies that prepare their financial statements using US GAAP must submit the statements contained within their Form 10-K annual and Form 10-Q quarterly filings in XBRL format. Among all US open data mandates for financial reporting, this one has had the broadest implications and generated the most controversy thus far. This study will secondarily explore a number of other open data mandates that are under consideration or have already been imposed.

This study will begin with an explanation of the technical issues and an examination of whether mandates to file financial information as open data, rather than unstructured documents, can be reconciled with free market ideology. Then, it will address the controversy over the SEC’s XBRL financial statement mandate and offer policy recommendations. The policy recommendations address two broad questions: First, what, if anything, should be done about the SEC’s current filing requirement? Second, what lessons can be learned from the SEC’s implementation of XBRL for other types of regulatory filings that may be converted to open data in the future?

After addressing the SEC’s XBRL financial statement mandate, this study will provide shorter summaries of the XBRL mandates for credit rating agency historical rating data, mutual fund reports, and Federal Deposit Insurance Corporation (FDIC) reports. Before concluding, it will list other types of SEC reports that may or will be covered by open data mandates in the near-to-intermediate term.

---

1 The author wishes to thank Megan Porzio and Christina Scheuer for their assistance.
4 Though both XML and XBRL transform document-based information into open data, they should not be considered equivalents or substitutable one for the other. See later “History” section for a discussion of the distinctions between these two data formats.
5 Open data is information that has been (1) standardized electronically and (2) published for easy access and reuse. See Alison Gill, Adam Hughes, and Hudson Hollister, State of the Union of Open Data, 2016 (November 2016), http://www.datafoundation.org/state-of-the-union-of-open-data-2016/, section 2.
Most people are familiar with bar codes on products, courier packages, and digital US postage stamps (see Figure 1). People no longer need to type a product price into a cash register or hand sort envelopes and packages, visually affirming sufficient postage. Machine-readable codes enable automation, thereby increasing efficiency and reducing errors.

The same benefits accrue for regulatory information when it is annotated with machine-readable tags that capture (in the SEC’s case, for instance) financial values and descriptive information about the values. XBRL is one of a number of formats that are used to report machine-readable, “self-describing” data. Unlike bar codes, XBRL data is represented in plain-text files (more on this below). And though the contents of these plain-text files can be read and understood by a technical expert, the information is meant to be generated and read by computer programs. Whether XBRL or another open data format is mandated, the intention is the same: data from the regulatory filing should be easily transmitted between systems, with perfect accuracy and without a loss in information. There should be no need to retype information or guess what a value means.

Computers can transform and present data to meet the wishes of individual data consumers (e.g., a table containing a filtered set of numbers with values rounded to the nearest thousand). In the specific case of SEC financial statement filings, the filer provides the data in XBRL format. Investors can then select portions of the data that interest them, transform and format it as they wish, and/or load it into a spreadsheet or database management system (DBMS) for quantitative analysis. With the standards established by this mandate, the investor has the ability to take the data beyond the filer’s presentation.

The XBRL data format is built upon a more general data format known as XML (eXtensible Markup Language). XBRL is primarily intended to support quantitative analysis. All XBRL files are XML files, but not vice versa. XML files are plain text files in which data is enclosed within a hierarchical series of tags. The following is a snippet of XML:

```xml
<company>
  <name>Apple</name>
  <year_founded>1975</year_founded>
  <ceo>Tim Cook</ceo>
</company>
```

6 A DBMS is like a collection of Excel worksheets, but databases have more capacity than spreadsheets and often impose more restrictions on data being entered (e.g. an individual column must consist entirely of text, numbers or dates).
This snippet of XML code provides the names, founding year, and Chief Executive Officers for two companies, Apple and Amazon. The opening and closing tags clearly tell people — and more importantly, tell a computer program — what each data point means, where it starts and where it ends. All XML tags are surrounded by angled brackets, and closing tags include a slash before the tag name.

Another benefit of this structure is that it facilitates automatic checking or validation. A validation program consults a list of valid tags — known as a taxonomy — to test the integrity of tagged values. A validation program can reveal cases in which a tag name is misspelled or the tag used for a value is incorrect (e.g., a tag is defined to describe stock shares but is incorrectly applied to a dollar value).

XBRL taxonomies can also describe relationships between tagged financial statement items. For example, a taxonomy can specify that the numeric value reported as “Total Assets” must also equal the sum of the numeric values reported for “Current Assets” and “Noncurrent Assets”; this relationship can then be checked by a validation program to make sure the value reported as “Total Assets” is consistent with its calculated value.7

Technically, a filing does not need to be provided in an open data format like XBRL in order for its contents to be transferred to a database or other software program. At the most crude level, the contents of printed regulatory filings can be manually typed into a database or software program. If a large volume of printed filings must be transferred, they can be scanned into a computer to produce image files. The resulting image files can be run through an Optical Character Recognition (OCR) program to derive text. The text files produced by the OCR process can then be parsed (programmatically analyzed) to identify field names, field values, and other contents such as reporting period, unit of currency, etc., and this parsed data may then be loaded into a database.

In theory, the three-step process of scanning, OCR, and parsing can be automated. However, each step involves complexities and the opportunity to introduce errors, thereby increasing the costs and/or compromising the value of the converted data. OCR software may, for example, confuse a lower-case “L” with the numeral “1” or confuse the numerals “3” and “8.” Likewise, paper forms may become jammed in a scanner just as they often are in copy machines.

In recent years, the paper form scenario has become increasingly irrelevant. Corporate regulatory returns are usually filed electronically, eliminating the need for scanning. Further, if these electronic filings take the form of text files or “searchable” PDFs, the OCR step is unnecessary because the electronic file already contains machine-readable text.

Thus, the first two steps of the conversion process can be eliminated without much burden on the filer of a regulatory report. As long as it is easy for the regulated entity to create and submit electronic filings to the regulator via email, file transfer protocol, or website upload, the filer is likely to find compliance to be either less costly than or roughly only as costly as filing on paper. In the SEC’s case, all regulatory filings must be uploaded to the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system and are usually submitted in text form.8

---


However, even with PDF or another electronic text format, the third step, parsing, remains. Text-based electronic filings must still be parsed before being loaded into a database because databases typically store discrete numeric and character data entries organized into rows and columns (i.e., database “tables”). Depending on the complexity of a regulatory filing, this parsing process may be very expensive and error prone.

Entries in a form that are concise and clearly identified are relatively easy to parse. For example, the form fragment below (see Figure 2) could be parsed with minimal programming effort:

![Company Registration Form]

<table>
<thead>
<tr>
<th>Company Registration Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name of Company</td>
</tr>
<tr>
<td>Year Founded:</td>
</tr>
<tr>
<td>Name of Chief Executive Officer:</td>
</tr>
</tbody>
</table>

The form contains clearly labeled spaces in which the filer enters a discrete and concise piece of information. Additionally, fields are in the same physical position on each form, providing another standard attribute that can be exploited by a parser.

But when a filing contains free-form text, parsing becomes much more of a challenge. Footnotes to financial statements provided in connection with Forms 10-Q and 10-K exemplify this scenario. US companies that hold retained earnings in overseas subsidiaries are required to disclose the amount of these earnings in a footnote. But since the exact position and format of the footnote is not consistent across filings, obtaining that data through parsing is costly and error-prone.

A similar concern applies to the balance sheet, income statement, and other “face financial” statements in a 10-K filing. Although these tables have more inherent structure than footnotes do, and may thus seem easier to assimilate automatically, they pose unique challenges for a parser. Row and column names have inconsistent meanings and terminology across filers (e.g., one company uses the label “Profit” and another “Profit/Loss” and a third “Profit/(Loss)”). Further, the meaning of each cell is defined by the combination of the row and column headers. Parsing software must determine which row header and which column header to associate with each cell. This task can become especially complex when column headers span multiple lines and columns are separated by minimal spacing.

Open data formats like XBRL greatly simplify the parsing process by requiring that each discrete data point be identified by opening and closing tags. In the case of financial statements contained within Form 10-K and 10-Q filings, every table entry and numerous footnote elements must be enclosed within appropriate XBRL tags. The XBRL formatted data, like all standards-based computer formats, prescribes exactly how to interpret the format; this is critical for the information to be machine-readable and communicated with fidelity. Standards can facilitate great efficiencies, economies of scale, and opportunities. For example, the rapid commercial growth of the Internet industry depended on computer data standards such as HTML and the Internet Protocol. Similarly, XBRL also is a standard — in this case a data format or syntax standard for financial information — and a taxonomy expressed in XBRL is an example of using one standard (i.e., XBRL) as a means to express another kind of standard (i.e., a data dictionary).

Next Steps: The Development of iXBRL

Inline XBRL, or simply iXBRL, is a newer variant. It expands the capabilities of XBRL by making it possible

---

9 Imagine an Internet where each web browser (e.g., Internet Explorer, FireFox, Safari, Chrome, etc.) used a proprietary format instead of the shared standard format of HTML. The consequence of multiple proprietary formats is that every web page would have to exist in each of these multiple formats, creating inefficiencies, raising the cost of providing information on the web, and frustrating users. By adopting a high-quality universal standard, the Internet industry avoided the inefficiencies that plague users of document editors who are forced to convert documents to use different editors (e.g., Microsoft Word, Apple Pages, WordPerfect, Wdesk, Google docs, etc.).
to view XBRL documents on any web browser, thus making the information much more accessible. The iXBRL format uses HTML, the normal language of the web, to allow files to be displayed by ordinary web browsers on any device. HTML files contain tags that control visual formatting when displayed in a web browser. The iXBRL standard adds XBRL tags to HTML files, thus allowing XBRL data to be viewed in a web browser in a human-friendly visual format (HTML) and to be extracted for use by XBRL processors. Embedded within these human-readable web pages, and wrapping individual facts, are iXBRL tags that instruct systems to mark up facts in XBRL. Web browsers disregard these sections of the text, but an XBRL processor can use the information to convert the HTML file into a traditional XBRL document. A single iXBRL file can support the attractive visual display properties of HTML while also offering the database integration benefits of XML and XBRL.

Using a single document to satisfy both data and visual requirements streamlines the reporting process because just one document needs to be prepared, instead of separate human-readable and machine-readable versions. It also eliminates the burden of keeping the content synchronized across multiple documents (see section below, “Multiple Documents Lead to Multiple Costs”).

---


BASIS FOR OPEN DATA REPORTING MANDATES: A FREE MARKET PERSPECTIVE

Before considering specific open data filing requirements, this study will address the theoretical case for these mandates from a free market standpoint. Under a complete laissez-faire system, businesses would not have to file any disclosure documents with government agencies. Given an ideal of voluntary disclosure, conservatives and libertarians tend to be skeptical of new government-mandated information gathering requirements. However, even within a free market economy, the use of standards and regulations facilitates trade by creating a more stable and transparent system.

However, even within a free market economy, the use of standards and regulations facilitates trade by creating a more stable and transparent system.

As this study later explores, compliance with open data filing requirements does impose some out-of-pocket costs on public companies. Some choose to pay a service provider to generate their XBRL disclosures, while others license software to generate their own XBRL disclosures. Further, some methods may create unnecessary time costs and pressures for the filer. As Shane Kovacs, CFO of PTC Therapeutics, noted in his testimony to Congress, “Outsourcing to an XBRL expert requires that the internal team complete the traditional filing statement with enough time to spare for the external contractor to complete the XBRL process before everything is due to the SEC.”

The fact that open data filings are being required in addition to, rather than instead of, legacy formatted financial statements adds to a business’s regulatory burden. Thus, it is unsurprising that some people in the financial reporting community have taken a dim view of XBRL. That said, a case for these requirements can be constructed in market-oriented terms.

Entities affected by open data financial filing requirements include corporations, limited liability companies, and other forms of legal entities. As such they are “creatures of the state,” enjoying special privileges such as limited liability. Since the ability to create a corporation — an artificial person — is a privilege conferred by the state, it is reasonable for the state to extract a public benefit from those entities in exchange for that special privilege; providing greater transparency through open data financial filings is one such public benefit. Allowing corporations

---


13 At the time of this writing, PTC Therapeutics has changed its SEC filings method to use a technology that unifies the preparation method of both their traditional filing and their XBRL filing, thus eliminating this impact on their filing schedule. Still, other public companies may continue using methods that present the challenges cited by Mr. Kovacs.

14 The SEC mandate only requires that the financial statements filed in XBRL format minimally cover the financial statements and certain schedules contained within Forms 10-K and 10-Q — not the entire Form 10-K or 10-Q. Thus, the potential for truly duplicate content is limited to the financial statements. For a detailed listing of what must be reported, see Securities and Exchange Commission, Final Rule: Interactive Data to Improve Financial Reporting. See also Government Publishing Office, Electronic Code of Federal Regulations, par. 232.405, last modified January 26, 2017, accessed January 31, 2017, http://www.ecfr.gov/cgi-bin/text-idx?node=17:3.0.1.1.14&rgn=div5#se17.3.232_1405.

to enjoy state-conferred privileges without any accountability is unjust. The opposite approach — removing both the limited liability privilege and its attendant regulatory burdens — seems politically infeasible.

Some scholars argue that the limited liability corporation would have developed in the absence of government intervention. While that might be the case, it is also likely that corporations operating under a laissez-faire system would face disclosure demands approximating government mandates. For example, even in the relatively unregulated world of ridesharing, prospective drivers cannot get access to the Lyft or Uber networks unless they submit to a comprehensive background check. The ridesharing companies demand this level of disclosure to minimize the number of rogue drivers who may inflict reputational damage (by harming their riders, for example). This might be called a “market discipline” arrangement because the regulations are created to protect the corporation, its investors, and its customers. Such regulations arise from a need created by the market itself.

Similarly, even in a laissez-faire system, corporations wishing to offer shares on exchanges and electronic trading platforms might be required by these marketplaces to provide periodic audited financial reports. If the NYSE or NASD allowed companies without publicly available audited financial statements to trade their shares, they could suffer reputational damage and a reduced volume of trading because buyers would worry that some listings were fraudulent. In markets driven by information, opacity and mistrust can be devastating. For example, one of the major causes of the 2008 financial crisis was the mistrust in a counterparty’s ability to fulfill the obligations of its agreements. Part of this mistrust arose because of the opacity of counterparty identity i.e., how can one trust an unknown counterparty? Market-making entities thus have an incentive to impose requirements on market participants, and even in a laissez-faire system, these requirements could resemble contemporary government regulations.

So, while it is possible to imagine corporations existing without government intervention, it is also possible to also imagine disclosure requirements arising under laissez-faire. That said, exchanges and trading platforms, operating in a competitive environment, would have a strong incentive to not overburden corporations, ensuring that marginal increases in disclosure burdens closely match the marginal improvement in investor satisfaction with the level of information provided. However, investors often measure satisfaction based on short-term results. On the other hand, the SEC has a mandate to provide investor protection. Investors generally equate protection with the avoidance of events such as financial fraud or market bubbles. These “hot” markets may enhance short-term results for investors but can lead to significant, sometimes catastrophic, market disruptions resulting in significant investor losses. Therefore, it is possible to make a free market argument in favor of mandates that require regulatory information to be filed as open data because if investors have access to more information, they can make more informed decisions and facilitate a healthier, more sustainable market.

---


18 The problem of counterparty identity, and alternatives to the over-reliance on a proprietary DUNS number particularly in the US, is being actively pursued by governments and non-governmental bodies.

19 As recent history has shown, in a bubble investors generally do not value protection with a full understanding of their self-interest (i.e., until it’s too late).
THE SEC’S XBRL REPORTING MANDATE FOR FINANCIAL STATEMENTS IN FORMS 10-K AND 10-Q

In the United States, the highest-profile mandate for open data regulatory reporting is the SEC’s requirement that publicly traded companies file their quarterly and annual financial statements in the XBRL format. After reviewing the history of this mandate, this study will survey the major criticisms of the mandate and then make some policy recommendations.

History of XBRL

In 1999, the American Institute of Certified Public Accountants (AICPA), along with development partners from the accounting, technical, and securities sectors, launched the XBRL initiative with the goal of providing a standard, XML-based language for digitizing business reports. The primary motivation for the creation of the XBRL format (as opposed to using native XML) arose from the need for machine-readable financial reports to do the following things:

1. describe how different tags relate to each other, sometimes in fairly complex ways, e.g., a mathematical relationship among tags or subcategories of a tag;

2. customize the taxonomy through the customization of properties of standard tags (e.g., labels) or the creation of new, entity-specific custom tags; and

3. allow the exchange and collation of structured data even when the taxonomy is customized.

The primary technical limitation of native XML is that it bundles together the definition of tags and the data being tagged. Therefore, a filer who creates a custom tag within XML can break comparability with other reports. By contrast, XBRL separates the data from the definition of the data structure; this allows the structures to be adjusted with minimal risk of “breaking” the data. The need to accommodate change in deference to filers was a compelling reason for the creation of XBRL.

By 2000, the AICPA had formed a standards group named XBRL International and began publishing specifications for the new language. Government mandates would follow in the early years of the new millennium; the FDIC’s 2002 decision to use XBRL for bank financial statement data (discussed later in this study) and China’s 2004 decision to require XBRL for corporate financial reporting were the most significant. China’s XBRL corporate reporting mandate was followed by similar requirements in India, Japan, Singapore, and South Korea. Today, XBRL is the open international standard for digital business reporting, managed by a global not-for-profit consortium, XBRL International (XII). XII is committed to improving reporting in the public interest. XBRL is used in more than 50 countries, and millions of XBRL documents are created every year, replacing older, paper-based reports with more useful, more effective, and more accurate digital reports.
XBRL is used in more than 50 countries, and millions of XBRL documents are created every year, replacing older, paper-based reports with more useful, more effective, and more accurate digital reports.

In 2005, then-SEC Chairman Chris Cox became interested in XBRL, placing it on a fast track toward adoption for corporate financial statements. Since the mid-1990s, public companies had been filing their financial statements with the SEC in plain text (ASCII) format, with HTML added as an option in 1999. Although HTML documents are easier to read in a web browser than plain text, they do not contain the descriptive tags required for database integration. Cox concluded that the prevailing system did not adequately serve twenty-first century investors. In a 2006 speech advocating for widespread adoption of XBRL, Cox said,

“What we need is something that will give individuals faster access to better information that they can easily use and understand. We need to make searches for information easier. It should be easy to call up information about any company you choose. You should be able to download it and use it in the personal software of your choice. And you then should be able to easily analyze and compare the data with the same information from other companies. We want to make the numbers derived from financial statements vastly more accurate. And we want to allow companies to communicate with investors on a constant basis. At a time when we have 24-hour news — and even 24-hour pizza delivery— why are we still living by the 10-K and the 10-Q?”

In 2008, XBRL US — the US jurisdiction of XII — published a taxonomy of XBRL elements based on US Generally Accepted Accounting Principles (GAAP) and called it the US GAAP Financial Reporting Taxonomy. At the end of that year, the SEC approved a final rule mandating public companies to file financial statements in XBRL using this taxonomy.

The new rule phased in the XBRL requirement over a four-year period. Large accelerated public filers worth over $5 billion were required to convert face financial statements to XBRL in 2009, and the mandate extended to other large accelerated filers in 2010 and then to all remaining public companies and foreign private issuers in 2011. All firms received an extra year to convert their footnotes into XBRL, so the mandate was fully implemented by June 2012.

---

30 Ibid. p. 42.
CURRENT DEBATE

In the next three sections, this study discusses three major criticisms of the SEC’s XBRL reporting mandate. The first criticism is that few investors use the published XBRL formatted filings. The second criticism is that these filings are costly and difficult to create. And the third criticism is that the SEC does not use the data internally and does not enforce data quality; therefore, the accuracy of XBRL data is suspect. These sections also show how the SEC, companies, and software vendors are addressing each of these problems and discuss how the processes can be improved.

Underutilization

Five years after the final rollout of public company XBRL financial statements, critics argued that investors’ use of this new type of data had been limited. In the 2015 House subcommittee hearing cited above, Shane Kovacs also testified that a number of Wall Street analysts he spoke with did not even know what XBRL was.31

An earlier 2012 Columbia Business School survey of investors and analysts found that less than 10% of respondents knew that they directly used XBRL data from the SEC or from third-party providers.32 However, as Dr. Suzanne Morsfield, one of the report’s authors, later noted, it’s important to understand the context of this statistic: “Most investors use structured data through data vendors, not directly,” and “the statistic in [our] study doesn’t reflect indirect use.”33 That is, many investors may be using XBRL structured data without realizing it. Further, those investors who were direct users cited the efficiency and effectiveness of XBRL data as a unique benefit motivating their use.34 Finally, nearly 90% of those who knew about XBRL-tagged filings said they would be interested in utilizing XBRL data.35 Thus, this low number of direct users from the 2012 report likely reflects an “innovator” or “early adopter” pattern.36

As noted at the beginning of this study, although the contents of XBRL files can be read and understood by technical experts, the contents are meant to be read by computer programs. Direct use of XBRL involves a substantial learning curve; thus, it is not surprising that recent interviews conducted by the author with XBRL industry participants revealed that direct investor use of XBRL has not increased markedly. As Professor Vasarhelyi and his colleagues observed, “Nonfinancial experts may feel far more comfortable using spreadsheets, a slow but reliable mainstay, over a relatively unknown piece of software accompanied by minimal support and literature. A lack of apparent demand may delay developers from investing in XBRL software, further compounding the issue.”37

However, it is now clear that use of XBRL data needs to be considered from an information supply chain perspective. In other words, a significant proportion of individual and institutional investors do not work...
directly with SEC filings. Rather, they rely on market data providers to gather the filings and organize them into a database. Key market data providers include Bloomberg, Thomson Reuters, and Morningstar. This study will now address how each of these leading data providers uses XBRL and then note several smaller companies who have a more singular focus on XBRL-formatted open data. An analysis of these users will allow us to better understand how XBRL data can be more fully utilized in the future.

**XBRL Data Use by Market Data Providers**

Bloomberg provides company financial data together with a large volume of other information on a proprietary platform known as Bloomberg Professional or the Bloomberg Terminal. This service has over 300,000 subscribers globally.38

According to Bloomberg’s Help Desk, the service compiles much of its US company data from XBRL filings.39 According to a member of Bloomberg’s management team,

> Bloomberg’s US XBRL coverage started with the third quarter of 2015 and we expect to process all 6,600 filers in 2016. Due to the complexity of US XBRL the data is not straight-through processed, but our XBRL engine serves it to an analyst for further processing.40

In this usage, “straight-through processing” refers to automatically loading XBRL data into a database and processing it without human intervention. Bloomberg states that it uses straight-through processing for XBRL financial filings in Japan, China, and South Korea, but does not do so in the US because of the burden of dealing with custom tags.41 (The issue of custom tags is covered in greater detail in the section below, “Complexity and the US GAAP Financial Reporting Taxonomy”).

Thomson Reuters, another major market data provider, has also processed XBRL filings for a number of years, and they provide customers with data derived from these filings on their Eikon42 service — a web-based competitor to the Bloomberg Terminal.

Morningstar does not currently use XBRL financial statement data but plans to do so in 2017. When XBRL data first became available, Morningstar analysts concluded that it contained too many errors to be considered reliable. However, they now believe that the data has improved and have started a project to integrate it onto their platform.43

A number of smaller companies also ingest and analyze XBRL filings for clients. For example, Calcbench, founded in 2011, claims to have a competitive advantage in financial data analysis because they “harness the XBRL data standard to collect amazingly rich and detailed information on every filer with the SEC.”44 According to CEO Pranav Ghai, the company offers a web-based interface for institutional investors at a cost of $400 per month. The interface allows users to view and analyze filings. Calcbench’s analysis features include comparisons of companies with peers, and leverages tagged data in each filer’s footnotes. For example, Calcbench may be used to create a list of companies retaining earnings offshore, showing the amount of profit each firm has yet to repatriate.45

---

39 Chat session with Bloomberg Help Desk, May 9, 2016.
40 Bloomberg statement on XBRL, e-mail message to the author, May 11, 2016.
41 Emil Efthimides (Project Manager at Bloomberg), in discussion with the author, January 25, 2017.
43 Jo Guo (Head of Fundamentals and Equity Data Operations, Morningstar), in discussion with the author.
45 Pranav Ghai (Co-Founder, Calcbench) in discussion with the author, March 11, 2016.
idaciti, another XBRL startup, offers strategic analysis and competitive research services to subscribers. Users can see how their company ranks against others based on a selection of Key Performance Indicators, which idaciti derives from XBRL filing data. According to co-founder Emily Haung, the company makes extensive use of XBRL data but keeps it hidden from its users. She said, “Our users don’t care about XBRL. They care about comparability.” The same can be said for users of web browsers: i.e., they do not care about HTML; they just want to view a web page.

Intrinio is another data aggregator startup in this analysis area. Like all data aggregators, the company translates the 10-Q and 10-K XBRL formatted filings into a normalized information model so that the XBRL data can be compared with financial data from other sources. The company provides this curated data in a format that is useful for creators of other applications, particularly mobile phone applications. In addition to appealing to individual and institutional investors and analysts, Intrinio is also trying to assist the growth in finance-oriented applications in the broader consumer markets.

Another data aggregator that uses XBRL filings is EDGAR Online, which provides data and analytics gathered from SEC filings. EDGAR Online’s I-Metrix product “is an interactive data and analytic tool that provides quick and accurate XBRL-tagged financial statement data via Microsoft Excel.” The company was acquired by RR Donnelley in 2012 for $70.5 million, but it is not clear how much of the firm’s acquisition value related to its XBRL expertise.

The second criticism of the XBRL mandate is that 10-Q and 10-K XBRL filings can be costly and difficult to create. If investors derive limited value from XBRL, these costs may not be justified. Some sources have suggested that implementation costs are substantial. Shane Kovacs testified that his company, PTC Therapeutics, spends nearly $50,000 per year on XBRL compliance. At the time of this writing, PTC Therapeutics has changed its SEC filings method to use a technology that unifies the preparation method of both their traditional filing and their XBRL filing, thus eliminating this impact on their filing schedule. Still, other public companies may continue using methods that present the challenges cited by Mr. Kovacs.

US XBRL filings are already informing how investors make decisions and how companies understand their relationships to their competitors. An analysis of these companies shows that market data providers understand the value of XBRL for market analysis and that US XBRL filings are already informing how investors make decisions and how companies understand their relationships to their competitors. Moreover, this data allows market data providers to provide a much better picture of market trends and the overall health of the market. The analysis of XBRL use by market data providers also shows us that many people who use this data are not aware that they are using XBRL-tagged data; they focus on the data itself rather than how it is structured or coded. Therefore, XBRL may be utilized much more widely than the initial numbers suggest.

Costs

The second criticism of the XBRL mandate is that 10-Q and 10-K XBRL filings can be costly and difficult to create. If investors derive limited value from XBRL, these costs may not be justified. Some sources have suggested that implementation costs are substantial. Shane Kovacs testified that his company, PTC Therapeutics, spends nearly $50,000 per year on XBRL compliance. A 2011 Financial Executives International report quoted annual compliance costs of up to $500,000 — a figure that was

47 Emily Huang (Co-Founder and CEO, idaciti), in discussion with the author, January 3, 2017.
48 Rachel Carpenter (Co-Founder and CEO, Intrinio), in discussion with the author, January 4, 2017.
51 For details regarding how the Taxonomy’s complexity plays into costs, see later section, “Complexity and the US GAAP Financial Reporting Taxonomy.”
52 Kovacs, Testimony, Legislative Proposals to Enhance Capital Formations and Reduce Regulatory Burdens, 9.
53 At the time of this writing, PTC Therapeutics has changed its SEC filings method to use a technology that unifies the preparation method of both their traditional filing and their XBRL filing, thus eliminating this impact on their filing schedule. Still, other public companies may continue using methods that present the challenges cited by Mr. Kovacs.
later repeated in a 2013 Wall Street Journal blog post and in a House committee report.

The $500,000 annual compliance cost estimate could not be verified for this study because the author was not available for comment. However, one interviewee speculated that the amount represented the acquisition price of software required to generate XBRL disclosures and perform other accounting functions. If this is the case, the annual estimate is misleading because the one-time software purchase cost should be allocated across the number of years during which it is used. Also, if the software in question is a general-purpose accounting package, its cost cannot be fully attributed to XBRL compliance.

A subsequent survey by the Financial Executives Research Foundation (FERF) in 2013 collected XBRL filing cost data from 382 public companies. FERF grouped the filers into four categories ranging from Large Accelerated Filers to Smaller Reporting Companies. Average outside services costs for their most recent annual filing ranged from $10,000 to $21,000, with the median costs ranging from $2,000 to $10,000. Since the sample of companies included in the survey was self-selecting (i.e., companies volunteered to participate in this survey), population statistics may vary from these sample results. Second, experience drives down costs, and therefore, in the three years that elapsed between the date of the FEI report and the FERF report, improvements in both software and filing practices have probably helped bring costs down.

A still more recent survey of XBRL filing agents complements the above filer survey. In 2015, the AICPA surveyed fourteen XBRL filing agents that provide XBRL services to 1,299 smaller public companies (i.e. companies with market capitalization of $75 million or less). The agents were asked to report the XBRL creation and filing fees they charged to these filers. In 69% of the cases, respondents charged public company clients $10,000 or less annually for XBRL services; 18% of the companies paid between $10,000 and $20,000 annually, only 8% paid more than $25,000 annually, and no one paid more than $50,000. Higher fees were often the result of rush charges arising from last minute changes to the company’s filings or from the complexity of the filings themselves, leading to additional XBRL creation services.

For all public companies, the median cost is undoubtedly higher than it is for smaller reporting companies. Emily Huang, formerly the CEO of XBRL tagging service Rivet Software (now operating under the name Certent) and more recently the co-founder of idaciti, recalls charging clients in the range of $10,000-$20,000 annually for XBRL services. The median costs for all public companies is likely to be in this range because larger public companies generally have more complex financial statements than their smaller counterparts, but the degree of additional complexity is not especially large. The length of the financial statement and footnotes is longer for larger companies than their smaller counterparts, but the difference is well below an order of magnitude. According to Travis Dyer et al., median Form 10-K length is 37,370 words, the 25th percentile length is 24,678 words, and the 75th percentile length is 55,852 words. They also find that company size is only one of several significant variables that explain Form 10-K length.

---

54 Lorraine Malonza, “‘XBRL-Speak’ No Easy Task for Filers, Revenue Recognition Standard to be Re-exposed,” Financial Executive, July 2011.
59 A data visualization company with a focus on public company reporting.
Multiple Documents Lead to Multiple Costs

Mike Starr, Vice President of Governmental and Regulatory Affairs at Workiva, a software as a service (SaaS) company and service provider, told the author that XBRL costs have been falling in recent years due to process improvements. After the mandate, many filings were handled by financial statement printers that first created traditional text-based filings and then converted them to XBRL — a two-step process. A more efficient approach is to use the same software platform to create a registrant’s financial statements and to tag the XBRL, simultaneously producing both text-based and XBRL formatted filings. This less costly method has been embraced by Workiva and several other service providers. The increasing innovation and competition by software companies for this market gives registrants an increasing range of options for preparing their filings. Thus, registrants are free to choose the option that fits their needs and budget.

On its face, requiring entities to produce the same disclosure in two different formats does not seem efficient. However, the burden may originate with the method rather than with the multiple formats. Most word processing programs can generate multiple formats (e.g., “Save As” a “docx” file for Microsoft Word documents, or as RTF, or as HTML, or as plain text, etc.). This generation in different formats is essentially without burden to the person producing the document. If one had to use separate editors and keep the contents between them synchronized manually, the burden would be significant. However, there are two ways to relieve this burden. The first is to standardize the expectation that disclosures be submitted in a single format capable of representing all the necessary information (such as iXBRL). The other option is to use a single editor that can produce multiple document formats via a “Save As...” function to support the multiple required formats (e.g., separate disclosure documents in HTML and XBRL formats).

In addition to the cost of preparing XBRL formatted filings, companies could face financial penalties if the XBRL disclosures contain inaccurate data. Inaccurate data could expose them to shareholder lawsuits because XBRL filings carry the same legal liability as the printed financial statements. When the SEC initially implemented the XBRL mandate, it deemed these new disclosures to be “furnished” rather than “filed” for an initial two-year period. Under Section 18 of the Exchange Act, companies that furnish materials to the SEC are not liable for misstatements or omissions, but they do have liability when the SEC deems the disclosure to be a “filing.” Now that the initial two-year period has ended for all public firms, the SEC considers all XBRL disclosures to be filed and thus subject to penalty for omissions and misstatements. As of this writing, the author of this study is not aware of any instance in which the SEC has penalized a company for improper XBRL filings, but given increased SEC attention to this issue, such sanctions are possible.

---

63 Mike Starr (Vice President, Governmental and Regulatory Affairs, Workiva), in discussion with the author, March 16, 2016.
64 iXBRL would be suitable for this task because it is capable of representing both the data component as XBRL and the visual appearance through HTML. See section above, “Next Steps: The Development of iXBRL.”
Costly Duplication of Information Already Available

Resistance to the SEC’s XBRL filing mandate has been particularly vigorous in the biotechnology industry, where companies have criticized it as both too costly and as irrelevant for this specific market. Because many public biotech firms are “pre-revenue,” investors are less interested in company financial statements and more concerned with the firm’s intellectual property. Such firms are best valued by estimating risk-adjusted future cash flows from their newly developed drugs.67 Charles Crain, Senior Manager of Tax & Financial Services Policy at the Biotechnology Industry Organization (BIO) trade association said that investors would rather see their money spent on lab work and drug trials instead of XBRL filings because it is only through the company’s research that investors can expect to see a return.68 Crain asserts that, for companies of this type, both traditional text financial filings and XBRL are redundant and useless.69

Crain supported the proposed Small Company Disclosure Simplification Act (HR 1965, 114th Congress), a bill that would have given companies with annual revenue below $250 million the option to not file XBRL disclosures.70 This measure was also supported by the US Chamber of Commerce, which stated,

> While XBRL was created in order to move away from a paper-based system of financial disclosures, it remains a work in progress and has experienced a number of growing pains. H.R. 1965 would allow [the] SEC to fix some of the deficiencies associated with XBRL. The optional exemption for EGCs (Emerging Growth Companies) and small issuers would allow boards and their shareholders the authority to decide whether using XBRL is in the best interest of the company.71

This view received support at the SEC Advisory Committee on Small and Emerging Companies. In a 2012 presentation to that committee, Paul Dorfman, Managing Director at NYSE Euronext, reported that members said,

> Eliminating the requirement for the XBRL would save time and money for smaller reporting companies. Smaller companies have less complicated financial statements and therefore, the value associated with being able to hyperlink to financial schedules and tagged footnotes is very low. Smaller reporting companies tend to use service providers that have long lead times associated with preparing XBRL files that do not work well with time constraints.72

The Committee recommended an XBRL exemption for smaller public companies in a 2013 memo to then SEC Chairperson Elisse Walter.73

---

68 Charles Crain, in discussion with the author, April 5, 2016.
69 As noted earlier, the XBRL formatted financial statements do not duplicate the printed versions of the Form 10-Q or Form 10-K. Thus, the potential for redundancy is limited.
70 Revenue may not be the most logical metric for determining who is subject to the mandate. Investors in a company with a public market value of $4 billion are risking $4 billion irrespective of whether that company has revenue of $1 or $250,000,001.
Yet, only four months later, the SEC Investor Advisory Committee, which represents a complementary constituency (i.e., investors), recommended both to expand the use of XBRL data and to take steps designed to reduce the costs of providing tagged data, particularly for smaller issuers and investors.\textsuperscript{74} Subsequent to both of these committee reports, SEC Chair Mary Jo White, in 2014 testimony to Congress\textsuperscript{75} expressed her opposition to the Small Company Disclosure Simplification Act (H.R. 4164), citing the value and efficiency of structured data and the critical role that data plays in the Commission’s work.

It’s important to put Dorfman’s 2012 statement, above, into a current context. First, his use of the word “hyperlink” mischaracterizes the purpose of XBRL formatted data. The term “hyperlink” describes a method for navigating across web pages. The primary goal of open data reporting, such as the SEC’s XBRL mandate, is to provide detailed meaning to the information in a regulatory report. Dorfman’s language understates the value of structured data reporting by assigning it such a small purpose. Second, as noted in the section above, experience and improvements in software have driven down the time and costs associated with preparing XBRL filings.

For consumers of this data, these smaller companies matter.

Lastly, it is worth noting that the universe of firms with annual revenue below $250 million is considerably larger than the universe of pre-revenue biotech firms. Mike Willis, the Assistant Director of the SEC’s Office of Structured Disclosure, estimates that 66% of US public companies reported revenues below $250 million in 2015.\textsuperscript{76} Further, investors appear to be interested in 10-K and 10-Q filings from smaller firms. According to Pranav Ghai, 85% of anonymous searches on Calcbench’s website in the fourth quarter of 2016 involved companies that were not in the S&P 500 index.\textsuperscript{77} For consumers of this data, these smaller companies matter.

**XBRL Use by the SEC**

The third and final criticism of the mandate is that the SEC does not use the data internally and does not enforce data quality; therefore, the accuracy of XBRL data is suspect. Ideally, when a regulator mandates a new format for open data reporting, it should review filings in that new format. In this case, the SEC not only mandates the disclosure of financial statements in XBRL for use by investors, but SEC analysts at its Division of Corporation Finance are also required to review every company’s disclosures at least once every three years: “In its filing reviews, the Division concentrates its resources on critical disclosures that appear to conflict with Commission rules or the applicable accounting standards and on disclosure that appears to be materially deficient in explanation or clarity.”\textsuperscript{78}

In 2013, Representative Darrell Issa, then Chair of the House Oversight Committee, sent a comment letter to SEC Chair Mary Jo White expressing surprise that the SEC was not using XBRL filings in its reviews, especially given the SEC’s knowledge of its benefits. Instead, the Corporate Finance Division staff were


\textsuperscript{76} Figure calculated by Pranav Ghai at Calcbench, January 18, 2017.

\textsuperscript{77} Ibid.

still utilizing market data providers (discussed above) and even “printouts, pencils, and calculators” for the financial statement data they examined.  

Representative Issa’s letter also referenced quality issues with XBRL filings, citing research by the industry consortium XBRL US. This group downloads all XBRL 10-K and 10-Q filings and runs automated checks on the information in these files. In 2013 they identified approximately 1.4 million possible errors. The 2013 letter from Representative Issa noted that the SEC had yet to publish a single enforcement action event arising from these possible errors. As of May 2015, the group had checked 128,193 filings filed by 10,491 companies and had identified over 5 million possible errors in these filings. Common errors identified included required values missing from the filing, negative values for an item that should always be positive, and duplicate items with inconsistent values.  

By failing to integrate XBRL filings in its work processes when the four-year phase-in for the mandate was complete, the SEC initially missed a valuable opportunity to review and improve the quality of these filings.

Subsequent to, but not necessarily related to, Representative Issa’s letter, the SEC increased its monitoring of XBRL filings. In July 2014, the Commission published a “Dear CFO letter” (a comment letter identifying a pervasive deficiency) noting that filings were missing one or more required calculation relationships. A calculation relationship asserts that one tagged value in a filing is a total of two or more other tagged values; thus, this assertion may be used to test the consistency of these mathematically related values.

The Commission’s Division of Economic and Risk Analysis (DERA) has also been evaluating XBRL filings to assess the overall use of custom tags — entity-specific tags that are not included in the standard US GAAP taxonomy. When filers use custom tags instead of standard tags provided by the Taxonomy, the filer’s data cannot be readily compared to that from other companies, and this also inhibits automated importing into databases (see Bloomberg example in the section above, “Underutilization”). There are legitimate needs for custom tags (see section below, “Complexity and the US GAAP Financial Reporting Taxonomy”); the issue is the creation of unnecessary ones. An example of an unnecessary custom tag, and one which is prohibited by the rules, is the creation of a custom tag when the filer’s disclosure requirements may be satisfied by simply customizing the label of a standard tag.

In a July 2014 Staff Observations report, the division reported a decline in the use of custom tags by large accelerated filers between 2009 and 2013, excluding from their sample certain industries which consistently

---

85 Ibid. See footnote 4.
have a higher than average custom tag rate,\textsuperscript{85} such as real estate and insurance companies. In a March 2016 report, the division noted a decline in the use of custom axis tags\textsuperscript{86} between 2013 and 2015.\textsuperscript{87} DERA also publishes a quarterly data set summarizing information from all XBRL financial statements it receives.\textsuperscript{88}

Other indications of the SEC’s increased interest in XBRL include:

- Its decision to license XBRL software. In late 2015, the Commission purchased more than 100 licenses for CalcBench, an XBRL analysis tool discussed above.\textsuperscript{89}
- Its release of an open source\textsuperscript{90} inline XBRL (i.e., iXBRL) viewer in EDGAR.
- Its incorporation of XBRL data into the Corporate Issuer Risk Assessment dashboard (CIRA). CIRA detects potential cases of fraudulent reporting by analyzing XBRL disclosures.\textsuperscript{91}
- Its development of a Financial Statement Query Viewer for internal use.\textsuperscript{92}

\textsuperscript{85} Ibid. See footnote 4.

\textsuperscript{86} In the US GAAP Financial Reporting Taxonomy there are standard tags for how to subdivide values by category (i.e., in XBRL these categories are represented by an “axis”), and the filer then makes custom tags to represent the entity-specific name for a member of that category (i.e., “Business Segment 1,” “Business Segment 2,” etc.).


\textsuperscript{90} The source code for the viewer, also called the EDGAR Renderer, can be found at https://github.com/Arelle/EdgarRenderer.


XBRL IN OTHER COUNTRIES’ CORPORATE FINANCIAL REPORTING REGIMES

Before offering policy recommendations, this study will briefly survey XBRL use within financial reporting regimes in countries other than the United States. In these cases, as in many others, experiences abroad can offer valuable lessons for US regulators. Let us again note that XBRL reporting in the SEC mandate represents only a small percentage of worldwide use and that the examples here represent only a portion of this worldwide use. For example, the Standard Business Reporting (SBR) program in Australia already results in more than 15,000,000 XBRL formatted reports annually to the Australian tax authorities. This dwarfs the 32,000 Forms 10-Q and 10-K filed annually with the SEC. The widespread use of XBRL for open data reporting shows us that XBRL can be successfully adopted on a very large scale, and it also suggests that the adoption of XBRL can facilitate US involvement in global markets.

The Standard Business Reporting (SBR) program in Australia already results in more than 15,000,000 XBRL formatted reports annually to the Australian tax authorities.

As noted in the “History of XBRL” section above, XBRL mandates in China, Japan, Singapore, and Korea preceded the SEC’s implementation. India has adopted widespread regulatory use of XBRL at its Ministry of Corporate Affairs (MCA), the Bombay Stock Exchange, and the Securities and Exchange Board of India. The MCA’s XBRL reporting requirements apply not only to public companies but also private companies with capitalization of Rs 100 crores or more (about $15 million USD) or paid up capital of Rs 5 crores or more (about $734,000 USD). Bloomberg processes XBRL filings from China, Japan, Korea, Taiwan, Israel, Brazil, and Mexico. Bloomberg also plans to begin capturing XBRL disclosures required in Peru and Colombia.

In the United Kingdom, every public and private company is required to submit its annual solo (i.e., unconsolidated, legal entity by legal entity) financial statements to the government’s Companies House, the UK business registrar. The same filings — i.e., full annual financial statements — must be submitted electronically to HMRC, the UK’s tax authority, as part of corporate tax filings. In terms of the voluntary business registrar filings, as of 2015, the majority of companies filed electronically, but larger firms have continued to make paper filings, in large part because transparency, particularly through machine-readable structured data, is not a high priority for

---

97 Bloomberg, e-mail to the author.
Those companies that do file open data have transitioned from XBRL to iXBRL.\(^9\)

It is worth noting that the United Kingdom requires companies to file earnings reports on a semi-annual basis, rather than the quarterly filing frequency mandated in the US. Semi-annual reporting is also required in Australia and New Zealand, while the European Union abolished quarterly financial statement filing requirements in 2013.\(^10\) Reduced reporting frequencies in the US would lower the cost of compliance, though it does come at a cost of transparency and utility because data diminishes in value as time passes. In April 2016, the SEC published a Concept Release in The Federal Register noting the EU’s decision and requesting comment on financial statement reporting frequencies.\(^11\)

The SBR program in Australia started in 2010 with the goal of streamlining the exchange of business information.\(^12\) Australia’s approach, focused on the creation of a taxonomy that could grow to encompass the standard business reporting already performed by conventional methods, rolled out incrementally. The SBR effort benefited from a focus on software vendor engagement to create “SBR-enabled” software; thus, the incorporation of XBRL was largely hidden from the users. Their incremental, software vendor engaged approach, combined with the benefits of a standard taxonomy and a standard data format with XBRL, has created success and momentum for the expansion of structured data reporting in Australia. SBR-enabled software has allowed Australian companies and regulators to automate reporting tasks that were previously manual, saving over $1 billion annually in compliance costs.\(^13\)

SBR-enabled software has allowed Australian companies and regulators to automate reporting tasks that were previously manual, saving over $1 billion annually in compliance costs.

---


COMPLEXITY AND THE US GAAP FINANCIAL REPORTING TAXONOMY

Some have claimed that the US GAAP Financial Reporting Taxonomy (“the Taxonomy”), which governs the SEC’s XBRL reporting mandate for public companies’ financial statements, is too complex, and argue that this complexity makes it difficult to create and consume the SEC’s XBRL formatted filings. This section will explain that most of its complexity arises not from the size or extensibility of the Taxonomy (via custom tags) but from a more fundamental yet correctable issue: the Taxonomy currently requires users to learn a new model for financial reporting rather than use the already-existing and familiar model of the US GAAP accounting standards.

As noted above, the problem with the Taxonomy is not its size, but rather challenges arising from unnecessary complexity. Though the Taxonomy does have more than 15,000 tags, in an automated world,\textsuperscript{105} the number of tags rarely matters. The ticketing system for a sports stadium is not necessarily more complex than a playhouse just because there are more seats in it. What would make it complex is a seat identification scheme that fails to match the actual seating plan. In other words, complexity arises when content does not match the description. Bringing this metaphor to the Taxonomy, the mismatch between the Taxonomy and what is required to be disclosed is the fundamental and, as will be explained below, correctable issue.

According to a statement by Mike Starr, former Chair of the XBRL US Data Quality Committee (DQC) and Vice President of Governmental and Regulatory Affairs at Workiva,

The US GAAP Financial Reporting Taxonomy was developed to support tagging printed financial statements rather than the required disclosures, introducing unnecessary variability into the Taxonomy. The US GAAP reporting standard for public companies\textsuperscript{106} provides the domain data standard for how the real or expected outcomes of similar events or transactions are to be disclosed in the primary financial statements\textsuperscript{107} and what additional information is disclosed in the notes to those statements about those events and transactions. This domain data standard and the required disclosures mean that companies provide the same or similar (in the case of entity specific disclosures) data for similar events or transactions. The Taxonomy needs to be revised to support the domain standard to eliminate the inconsistencies in the XBRL data.\textsuperscript{108}

What this means is that in order to comply with US reporting requirements, registrants must disclose the same or similar information for similar events or transactions. That’s how comparability is made possible. However, describing and formatting these values in printed financial statements is to a large extent discretionary, and generally depends on factors such as personal preference, an attempt to communicate more information, or simply historical


\textsuperscript{108} Mike Starr, e-mail message to the author, January 8, 2017.
practice i.e., “That’s how we described it in last quarter’s filing.” This discretion, while useful for a company pitching its financials as it sees best, unfortunately creates obstacles to comparing financials across companies. For printed filings it requires the reader, in some instances, to interpret facts’ meanings accurately to determine comparability; in machine-readable documents, this discretion results in unnecessary extensions and significantly hampers automated processing of the data.

A better practice, and one currently under discussion with the SEC staff, would be to base the Taxonomy directly on the US GAAP disclosure requirements for public companies and other widely-recognized disclosures. The US GAAP reporting standard has tremendous breadth, depth, and maturity. The Taxonomy would still be large — perhaps appropriately large — in order to achieve close alignment with what financial reporting staff already understand: the required disclosures. Instead, the approach used to create the Taxonomy, a decade ago, was based on the structure of printed financial statements. This unfortunately incorporated cross-company variations into the Taxonomy. These variations are the root cause of the inconsistencies in tagging data for similar events or transactions, and they hamper comparability.

Basing the Taxonomy on the structure of the printed financial statements created unnecessary complexity for XBRL filings based on the US GAAP Financial Reporting Taxonomy. Starr notes, “The Taxonomy is currently based on paper filings, and has led people to doubt the value of XBRL. It should have been (and still could be) the gold standard for how to build a XBRL taxonomy for financial reporting.” While it’s still a work in progress, the lessons from this experience will hopefully inform future revisions to the Taxonomy.

A taxonomy based on a rule-based accounting standard, such as US GAAP, is more likely to enable robust structured data comparability than one based on principles-based reporting, such as the International Financial Reporting Standard (IFRS). This is not to imply that principles-based reporting is not an appropriate basis for a financial reporting standard, but rather to say that it is more difficult to codify in a taxonomy while retaining consistency in the choice of tags. The difficulty arises when building a taxonomy solely based on principles-based financial reporting standards when there might not be, in some cases, specific disclosure requirements that capture the intent of the principle. Comparability requires consistent tagging. It is more likely that the IFRS taxonomy might need to move closer to the US GAAP Financial Reporting Taxonomy than vice versa.

Some critics argue that extensibility makes the Taxonomy unnecessarily complex. Broadly speaking, the SEC mandate allows filers to create custom tags. While this practice is common in the US, it is relatively rare in Japan and Taiwan. In China, XBRL submission tools provided by the Shenzhen Stock Exchange and the Shanghai Stock Exchange prevent filers from using extensions in most cases. While it may improve comparability

109 Starr, e-mail.
110 In cases where US GAAP does not provide guidance for disclosing an event or a transaction, there may be non-authoritative sources for the disclosure such as widely recognized and prevalent general or industry practices. In addition, the printed financial statements may include standard common presentations of information that are included in the Taxonomy because such presentations are widely recognized.
111 These “personalization” obstacles facilitate a market for market data providers for both paper and XBRL filings, thus keeping costs high to investors who wish to access curated data. However, improved data aggregation resulting from a Taxonomy based on the domain data standard will improve data quality, reduce costs for investors, and reduce barriers to entry into the data aggregation business.
112 Starr, e-mail.
113 Admittedly, this assertion that US GAAP is rule-based and IFRS is principle-based is not as black-and-white as it sounds. Still, the point remains: rule-based standards are easier to represent in software code than principle-based standards.
to prohibit custom tags, the decision to limit the use of extensions was considered and rejected in the SEC rule-making process. Extensions are permitted in the original mandate because they allow filers to add interpretive meaning to their financial data. Prior to the XBRL mandate, market data providers, who had only the printed financial statements to work with, had to make these interpretive decisions in order to transfer the contents of printed financial reports to the provider’s database(s). With this mandate and the availability of the Taxonomy, it is the filer who is now responsible for applying this judgment. According to the SEC Final Rule, the risk of interpretability of reported financial data already exists in the current data aggregation process. According to current practices, financial data service providers manually key financial information into a format that allows aggregation so that they can resell it to investors. As a result, the data service provider makes interpretive decisions on how to aggregate reported financial items so that they can be compared across all companies. This is done so that a subscriber of the commercial product offered by a data service provider can expect consistent interpretation of the reported financial items, allowing comparability in the same way that it is intended with interactive data. Hence, from one perspective, adoption of interactive data will shift the burden of making the interpretive decision on how to label a financial item from financial service providers to the companies making the filings. To the extent that the company is better able to classify financial data for comparability to other companies through interactive data tagging than a financial data service provider who manually keys and classifies financial data from standard paper-based filings, then interpretability of reported financial data should not worsen with adoption of interactive data reporting.\(^\text{116}\)

This excerpt from the XBRL mandate makes it clear that the SEC believes the filer is in a better position to make “interpretive decisions” regarding the filer’s financials (compared to financial printers and data aggregators). This is exactly where the discretion belongs.

It is also clear from the above statement that the intent of the Final Rule is to limit extensions and instead use standard elements as much as possible, though experience has shown that filers have not held themselves to this standard (see section above, “XBRL Use by the SEC”). However, a filer’s ability to create a custom tag provides critical utility because there are entity-specific disclosures that require it. For example, a filer in their segment reporting disclosure might report a unique source of revenue for which there is no standard element. The rule permits custom tags for these cases, and the US GAAP Financial Reporting Taxonomy ably supports it.\(^\text{117}\)


\(^{117}\) In the US GAAP Financial Reporting Taxonomy there are standard tags for how to subdivide values by category (i.e., in XBRL these categories are represented by an “axis”), and the filer then makes custom tags to represent the entity-specific name for a member of that category (i.e., “Business Segment 1,” “Business Segment 2,” etc.).
INDUSTRY STEPS TOWARD IMPROVING XBRL FILINGS

In June 2015, XBRL US, the AICPA, and an alliance of XBRL software and service providers serving a majority of the SEC-registered public companies for the preparation and filing of their financial and business data formed the XBRL US Center for Data Quality (“the Center”). Driven by a shared mission to improve the usability of XBRL data, the Center funds the Data Quality Committee (“the Committee”), whose membership consists of service providers, data consumers, preparers, and other market participants. The Committee’s meetings are attended by observers from the FASB and the IASB and open to the public. Committee members meet regularly with the staff of the SEC. To address concerns about the quality and usability of XBRL financial data filed with the SEC, the Committee develops appropriate guidance and validation standards (freely available to all public companies) to prevent or detect inconsistencies and errors in the XBRL data. Since November 2015, the Committee has issued over a dozen Validation Rules and Prescriptive Guidance documents.

Mike Starr, former Chair of the Committee, stated that “the mission of the XBRL US Data Quality Committee is to improve the usability of XBRL data that is filed with the SEC. To accomplish this mission, the Committee’s two objectives are first to ensure consistency of how the data is tagged and modeled, and, therefore, the comparability of the data; and second, to simplify how the data is tagged.” According to Starr, the Committee’s goals are consistent with and informed by the SEC’s stated intent to enable automated processing and analysis of financial information by requiring the use of a standard set of XBRL tags and limiting the use of custom tags.

The Committee aims to eliminate unnecessary and redundant elements from the US GAAP taxonomy, standardize tags and the way in which disclosure requirements are represented, and reduce the use of custom tags. In a recent release, the Committee forbade the use of custom extensions for Fair Value Measurement Frequencies and Bases, while limiting extension use for many other axis types. (An axis is a classification system, like a list of currencies or countries.)

Another industry step towards improving XBRL reporting occurred on June 13, 2016, when the SEC upgraded the EDGAR system to accept iXBRL filings and began allowing companies to voluntarily file iXBRL disclosures through March 2020. The first voluntary iXBRL filing occurred on July 1, 2016. The use of iXBRL expanded in subsequent quarters. On March 1, 2017, the SEC proposed a rule that would require the use of iXBRL for public company reporting, replacing the individual XBRL, HTML and ASCII files. This is a welcome first step toward replacing what are currently separate text and XBRL submissions with a single iXBRL filing. This proposed rule will solve a

---

120 Securities and Exchange Commission, Final Rule: Interactive Data to Improve Financial Reporting, 104. In order to promote comparability across companies, the new rules, as proposed, will limit the use of extensions to circumstances where the appropriate financial statement element does not exist in the standard list of tags. The new rules also require that wherever possible and when a standard element is appropriate, preparers change the label for a financial statement element that exists in the standard list of tags, instead of creating a new customized tag.
long-standing misuse of XBRL by public companies submitting XBRL filings to the SEC — the attempt to (mis)use the XBRL instructions in order to coerce XBRL viewing software to render a particular visual appearance. As described above, it is the iXBRL format that should be used if such control over appearance is desired, as iXBRL leverages the normal language of the web, HTML, to control visual presentation. Embedded within these human-readable HTML web pages, and wrapping individual facts, are iXBRL tags that instruct systems to mark up facts in XBRL. Thus, a single file can support the attractive visual display properties of HTML while also offering the database integration benefits of XBRL.

Regarding other industry improvement efforts, the Financial Accounting Standards Board (FASB), which submits annual US GAAP taxonomy updates to the SEC, has been working to remove unused, misapplied, and redundant tags. As evidence, FASB deprecated 441 tags in the 2015 taxonomy update and deprecated an additional 681 tags in the 2016 update.\textsuperscript{126}

RECOMMENDATIONS FOR IMPROVING THE SEC’S XBRL REPORTING MANDATE FOR FINANCIAL STATEMENTS IN FORMS 10-K AND 10-Q

This section first addresses the trends that are driving the move from document-based regulatory reporting to open data regulatory reporting. It then presents specific recommendations on how to make the SEC’s XBRL reporting mandate more efficient and effective.

The “Introduction” mentioned the possibility of abolishing the SEC’s XBRL mandate. However, this study will not recommend that option because it runs counter to the direction in which most of the world’s regulators and industries are going: toward increasing open data reporting, both business-to-government and business-to-business. The first trend driving the adoption of open data reporting is the predominance of data in management. While individual persons see themselves as managing information with documents, large organizations manage information as data. After Oracle Corporation pioneered relational DBMS technology in the late 1970s, the database industry grew to $33 billion in annual sales in 2015, with International Data Corporation (IDC) predicting annual growth of 6.6% through 2019.\(^{127}\)

While individual persons see themselves as managing information with documents, large organizations manage information as data.

The second trend arises from increasing expectations regarding the utility of data that are forcing the transition from plain text and semi-structured data to consistently-applied electronic tags, such as with XBRL and other types of data structures.\(^ {128}\) Data users expect increased reliability, the elimination of ambiguity of the meanings and types of numeric values (e.g., unit of measure, precision), data collaboration, audit tracking, etc. Given this trend, it seems inappropriate for regulators to resist the movement of regulatory reporting from documents to open data, especially by businesses who are themselves engaged in this migration for their internal information management.

Regulatory regimes worldwide are moving from document-based reporting to open data reporting, a move often called “disclosure modernization.” In some cases they are forgoing document-based reporting altogether. For example, the Federal Energy Regulatory Commission (FERC) is exploring a transition from document-based reporting to structured data collection as it modernizes its systems. Under this transition, FERC would consume the data and then make the data publicly accessible as both a structured data set and as a visually-oriented document equivalent (currently as a PDF).\(^ {129}\)

To participate in this trend towards open data, regulators will need to expand and improve their filing and reporting systems. In the following pages, this study presents four specific policy and technology use recommendations. These recommendations focus primarily on the SEC’s XBRL reporting mandate for financial statements in Forms 10-K and 10-Q, but could also be adapted by other regulators considering, or expanding, open data reporting mandates.


\(^{128}\) Relational DBMS prioritize data organization around storage and retrieval. Structured data, as used in this report, prioritize data organization around what the data means.

Align the US GAAP Taxonomy with compliance requirements

The first recommendation applies to the US GAAP Taxonomy and related guidance. In order to ensure that the financial data is tagged in a uniform, consistent manner, the following actions should be taken:

- The Taxonomy should be revised to establish an explicit connection between specific disclosure requirements for public companies under US GAAP and specific standard tags of the Taxonomy.

- The Taxonomy’s design should be revised to represent a data model which aligns with the model of the disclosure requirements. This aligns the shape of the two different methods for documenting the disclosure requirements.

- Policies and guidance for using the Taxonomy should be revised to restrict the use of custom tags.

- Preparers should select standard elements based on the required disclosures under US GAAP rather than a filing’s organization and appearance as a paper report.

- When the use of a custom tag is permissible, it should be connected to a standard tag of which the custom tag is a subset.

These recommendations will help realize a future version of the US GAAP Taxonomy that is an accurate, complete, and consistent reflection of the reporting requirements under US GAAP and widely-recognized common disclosures.

Automate data quality tests

The second recommendation is the automation of data quality tests. The best practice is simultaneous development of a taxonomy and its validation rules; a taxonomy should thus be designed with the goal of enabling the creation of automated validation rules. For example, if one wanted to make sure that certain values are positive numbers, one would need to make sure that the data quality tests can determine which values are supposed to be greater than zero. If one wanted data quality tests to confirm that only like-kind numbers are added together (e.g., all US dollars, all Euro, etc.) one would need to make sure that the data quality tests can determine the unit of measure for each value. Data quality automation needs to be built in to the Taxonomy, not layered on as an afterthought.

A complementary recommendation arises from the fact, noted earlier, that XBRL US found over five million potential errors in US XBRL filings through early 2015. Most of these errors could have been avoided if XBRL filings had to pass through an XBRL validation tool before being posted on the SEC’s EDGAR system. Unfortunately, in the case of US GAAP Financial Reporting Taxonomy, validation rules were not released together with the Taxonomy. Retrospectively, in the case of SEC, the workflow brought on by the XBRL mandate should have been developed and tested within the Division of Corporation Finance, and Division analysts should have immediately begun reviewing XBRL filings when registrants started submitting them in 2009. That way there would have been immediate feedback and corrective actions when widespread errors in the XBRL documents first appeared.

As of early 2017, SEC’s EDGAR Filer Manual listed several dozen rules for the content of XBRL documents, but only a subset of these are enforced.
automatically. As noted above, the DQC is developing open source validation rules\textsuperscript{132} to complement what the SEC currently has in place. If the SEC agrees that the DQC’s open source rules are reasonable, then the recommendation is that the SEC could and should incorporate them into the SEC’s EDGARLink Online Submission Tool. In the long run, regulators should work closely with a broad range of users when implementing new requirements.

The first two objectives of aligning and automating, above, are mutually reinforcing. This alignment of the Taxonomy with the disclosure standards under US GAAP can clarify guidance and simplify the creation of automated data quality tests. The net result is consistency in the tagging of XBRL data and cross-company comparisons.

Phase in iXBRL and phase out text-based and traditional XBRL filings

The next recommendation addresses financial report formats specifically: the SEC should adopt iXBRL and eliminate both text-based filings (i.e., HTML) and traditional XBRL filings. Because iXBRL addresses both display and data integration needs, it can be the sole format required from filers. The SEC is on its way to requiring this. As noted above, on March 1, 2017, the SEC proposed a rule that would require the use of iXBRL for public company reporting, replacing the individual XBRL, HTML and ASCII files.

On March 1, 2017, the Commission also proposed a phase-in period based on company size.\textsuperscript{133} This phase-in period is curious because the effort for compliance is not related to company size but rather the choice of a company’s vendor. Instead, the SEC should decide on the critical mass of vendors needed to assure the availability of compliance options for filers and set the date accordingly. These options, and thus the date, may already be close at hand; a survey of vendors would inform this decision and likely accelerate the mandate. In any event, the SEC should finalize its iXBRL proposal as a rule as soon as possible.

Additionally, other government regulators moving from documents to open data reporting should learn from the SEC and its constituencies’ experience in the implementation of its XBRL mandate. The SEC’s experiences suggest that future open data projects should prioritize iterative development. These efforts have a lot of moving parts, and so it is unrealistic to think of a transition as pulling a “big switch.” Rather, these projects are better done in relatively short cycles of effort: measuring the results, deciding on revisions, implementing the revisions, and then moving to the next release cycle. The SEC has been slow to revise, and that has been to the detriment of the Commission, the filers, and the investors.

Design taxonomies for reuse

Investors and analysts assume comparability across financial reports. There is no question that top-line revenue should be comparable across companies. This is why the US GAAP Taxonomy must be a complete and consistent reflection of the reporting requirements under US GAAP and widely-recognized common disclosures. This connection to the reporting requirements is what makes a taxonomy usable by a single reporting entity and more critically for all of the reporting entities.

Another type of reuse to consider is the reuse of the taxonomy by another taxonomy. For example, a regulatory agency may wish to capture information that relates financial activities to regulated activities, such as linking the use of assets (from a balance sheet) to


\textsuperscript{133} See III(C)3 of Securities and Exchange Commission, Proposed Rule: Inline XBRL Filing of Tagged Data.
specific activities of interest to the regulator (e.g., energy generation). Rather than create a new balance sheet and tags just for this, the regulator could instead reuse the balance sheet tags from the US GAAP Taxonomy, making an extension taxonomy that creates only those tags necessary for that regulator’s specific interests.

Taxonomy creators should expect that their work will have a life beyond these specific instances because a taxonomy is public information. (In contrast, most database designs are kept private.) Like a lobster, a taxonomy’s structure is on the outside (i.e., an exoskeleton). The structure is there for all to see.

Lastly, and with an eye to reuse, taxonomy creators should implement them using a computer language that reflects fundamental accounting models. XBRL is an ideal choice for these efforts because it is designed around fundamental accounting concepts, and it reflects the accumulated knowledge of accounting professionals. A low-level language (compared to XBRL), such as XML, may be extended to create these fundamental components for use in creating a financial reporting taxonomy, but since this work has already been accomplished with XBRL it may be better to forgo the simplicity of XML for the maturity and consistency of XBRL.
OTHER XBRL REPORTING MANDATES

This section explores other instances in which XBRL reporting mandates have been used to transform financial regulatory filings into open data; an analysis of these other mandates can show the benefits and challenges of the XBRL format. The first mandate was imposed by bank regulators, including the Federal Deposit Insurance Corporation (FDIC). The latter two mandates were implemented by the SEC. These additional cases add to our knowledge of what works and what does not work in the realm of open data financial regulatory reporting mandates.

FDIC

As mentioned in the “History of XBRL” section of this study, the FDIC implemented XBRL for reporting before the SEC. The FDIC replaced a form previously submitted via paper and PDF. The initiative was coordinated by an inter-agency working group known as the Federal Financial Institutions Examination Council (FFIEC), which also serves the Federal Reserve Board (FRB), the Office of the Comptroller of the Currency (OCC), the National Credit Union Administration (NCUA), and the Consumer Financial Protection Bureau (CFPB).

FFIEC formed a working group that selected a software vendor called Unisys Corporation to perform the implementation and then oversaw its work. FFIEC also administered a cost-sharing relationship between the FDIC and two other participating agencies (the FRB and OCC) that also used the affected disclosures.

The project resulted in the establishment of an FFIEC Central Data Repository (CDR) which contains quarterly reports of condition and income — known as Call Reports — filed by FDIC-insured banks. The CDR replaced a variety of legacy reporting formats with a consistent set of XBRL-formatted open data disclosures. The Call Reports include about 1,200 financial statement items. By 2005, about 8,000 institutions were filing them.

According to then FDIC Vice Chairman Martin Gruenberg, the switch to XBRL resulted in faster and more accurate bank filings, as well as increased productivity on the part of agency analysts. Individual FDIC analysts increased their caseloads from 450-500 banks to 550-600 banks each. Analysts completed their quarterly reviews seven days faster than they did prior to the implementation. Also, because most of the call report data is publicly available, the speed and accuracy improvements also benefited independent analysts monitoring bank performance.

Implementation of the CDR with XBRL was relatively smooth. When the switchover occurred, five external vendors were able to produce the XBRL-formatted filings and could thus help banks meet the new requirements. Further, FFIEC and member agencies...
prepared themselves for the switchover. According to the FFIEC Annual Report,

The Change Management Focus Group was active in training agency staff on the functionality of the CDR, preparing informational materials and press releases for the CDR web site and the media, communicating the project’s status to stakeholders, and ensuring the validation of metadata prior to implementation.\footnote{Federal Financial Institutions Examination Council, FFIEC Annual Report 2005 (March 30, 2006), p. 12, accessed January 24, 2017, https://www.ffiec.gov/PDF/annrpt05.pdf.}

It is worth noting that, although the FDIC’s number of filers and frequency of filings were comparable to the SEC’s scenario, the Call Report implementation was much easier than the conversion of 10-K and 10-Q reports. The FDIC Call Report taxonomy was developed based on the call report form, which contained discrete fields and restricted the addition of new fields.

Because these disclosure requirements were already rigid, the taxonomy developed for the new XBRL-formatted call report was easy to create and perfectly matched the disclosure requirements. Contrast this with the flexible nature of the Form 10-K and 10-Q reports (described above in the section “Complexity and the US GAAP Financial Reporting Taxonomy”); those reports permit entity-specific rearrangement of content, whereas the FDIC Call Report does not.

To date, the FDIC’s implementation of XBRL is the easiest US example to deem a success. The creation of the FDIC’s taxonomy was closely based on existing disclosure requirements, and those disclosure requirements were already unambiguous and rigid (i.e., the regulatory requirement for financial institutions to disclose certain prescribed financial information). The SEC’s implementation cannot yet be deemed a success, however, and the root cause is that the creation of the Taxonomy focused on supporting the organization and approximating the look and feel of printed financial statements, rather than on matching the underlying disclosure requirements in US GAAP. See the earlier section “Recommendations for Improving the SEC’s XBRL Reporting Mandate for Financial Statements in Forms 10-K and 10-Q” for details on this “root cause.”

The FFIEC’s implementation, with its thorough planning and change management, is a good model for other regulators with the same requirements. Thanks to the successful implementation of this project, the general public now has access to ten years of machine-readable financial data for FDIC insured banks.\footnote{This data can be searched and downloaded at Federal Financial Institutions Examination Council, FFIEC Central Data Repository’s Public Data Distribution web site, accessed January 24, 2017, https://cdr.ffiec.gov/public/.}

Mutual Funds

The SEC requires mutual funds to report historical return data and fee information in XBRL format. This data, known as “risk/return summary information,” is included with forms 485BPOS and 497, through which mutual funds file their prospectuses and prospectus supplements. In addition to submitting the XBRL data to EDGAR, fund companies must also post the data on their websites. The SEC applied an XBRL requirement to all mutual funds on January 1, 2011. As is the case with 10-K and 10-Q submissions, filers must submit information in both the XBRL and the legacy text versions.\footnote{Securities and Exchange Commission, Final Rule: Interactive Data for Mutual Fund Risk/Return Summary, Release 33-9006 (2009), 1, accessed January 24, 2017, https://www.sec.gov/rules/final/2009/33-9006.pdf.}
The impact of this SEC mandate has been much smaller than that for the XBRL public company financial statement requirement. According to the Investment Company Institute, 867 companies who sponsored US mutual funds at the end of 2014 have submitted XBRL reports, much fewer than the public companies that have submitted XBRL data with their 10-K and 10-Q reports (although it should be noted that many of these 867 fund companies must make dozens of 485BPOS and 497 filings annually). Further, the scope of information that has to be tagged within each fund’s disclosure is relatively small — covering only a few schedules and related notes in each fund’s prospectus. As a result, the mutual fund taxonomy includes far fewer elements than the US GAAP financial reporting taxonomy, making it easier to use.

The more limited impact of the mutual fund reporting mandate probably accounts for the relative lack of controversy. Although the Investment Company Institute submitted a comment letter asking for a delay of the mandate when the SEC first proposed the regulation, the author did not find calls to repeal the mandate once it had been imposed.

On the other hand, the XBRL risk/return data may not be receiving substantial use. Morningstar, a market data provider that offers detailed mutual fund information on its website, in correspondence with the author, said they do not use the disclosures. The company ignores the files due to their lack of timeliness. SEC rules require funds to provide the XBRL risk/return files as many as 15 business days after providing the equivalent text disclosure. Because Morningstar requires faster turnaround, it continues to use the text version of the forms, using software to extract the data it needs.

So while the risk/return XBRL mandate is not particularly onerous, the disparity in filing deadlines (i.e., the XBRL version can be filed up to 15 days after the paper filing) limits its utility. The SEC should move forward with its proposed rule of March 1, 2017 to replace the text-based risk/return disclosure with one in iXBRL, and move up the filing deadline for the iXBRL version to be submitted no later than the effective date of those filings.

Credit Ratings

SEC Rule 17(g)-7b requires Nationally Recognized Statistical Rating Organizations (NRSROs) to post XBRL files containing credit rating histories on their websites. The files, which are to be updated monthly, show the names and identifiers of issuers and securities along with the letter ratings they were assigned. Each time a rating changes, a new XBRL fact containing the date of the rating action and the new rating is added to the file.

In theory, these files should be quite useful to academics studying credit ratings. At a recent meeting of the American Society of Public Administration, the author asked a group of public finance academics interested in credit ratings whether they were aware of these XBRL disclosures. None of the attendees knew about them. In addition, one smaller credit rating agency told the SEC in 2014, that, on average, its credit rating history was being downloaded only once each month. These anecdotal observations suggest that the rating history disclosures are not widely used. Three aspects of the XBRL credit ratings implementation limit the value of these disclosures: timeliness, location, and data format.

---

145 Both schemas can be downloaded from the SEC’s Standard Taxonomy Page at https://www.sec.gov/info/edgar/edgartaxonomies.shtml. When reviewed on January 24, 2017, the compressed file containing the latest US GAAP taxonomy was 5274 kb, whereas the latest Mutual Fund Risk/Return Summary compressed file was only 55kb.
147 Securities and Exchange Commission, Interactive Data for Mutual Fund Risk/Return Summary: Final Rule. 1.
148 Sagar Patel (Morningstar), in correspondence with the author.
Rule 17(g)-7 allows rating agencies to post rating changes on a delayed basis. If a rating was paid for by a debt issuer, it need not be included in the XBRL disclosure for twelve months after the rating was issued. This twelve-month grace period also applies to upgrades and downgrades of issuer-paid ratings. In the case of investor-paid ratings, the delay is twenty-four months. These long delays greatly reduce the value of these files to fixed-income investors who need access to current ratings as part of their analysis. Instead, investors obtain individual ratings from credit rating agency websites or pay subscription fees to the rating agencies to obtain more current data. In its 2014 NRSRO rulemaking release, the SEC noted that longer grace periods “reduce potential losses experienced by NRSROs in revenues from services that include access to their credit ratings and/or rating histories.”

Second, the rating history disclosures are not stored in a central repository like the SEC’s EDGAR system or the FFIEC’s Central Data Repository discussed above. Instead, each rating agency posts XBRL filings on its own website. Since there are ten NRSROs, a user would need to visit ten different websites to gather all the rating history disclosures. (This problem is somewhat mitigated by the fact that 96% of the credit ratings are issued by the top three NRSROs: S&P, Moody’s and Fitch.) Further, at least some rating agencies require users to obtain an ID and password before downloading the disclosures, and the terms of access may restrict or prevent the (re)publication of the data or of derivative analysis. Such restrictions may deter market data firms from using the data. Government repositories of regulatory disclosure, like EDGAR, do not have a login requirement or license restrictions.

Finally, there are technical barriers limiting the data’s use. The first barrier is simply the great quantity of files to work with. Though a person wishing to aggregate the data for analysis would only need to visit ten different NRSROs, each rated financial instrument has its own file set, resulting in the need to manage thousands of files. The design of the XBRL taxonomy further complicates the use of these files. The latest ratings XBRL taxonomy contains approximately 50 elements that could potentially describe any given rating. These elements may pertain to the rating itself, the instrument being rated, or the obligor. A taxonomy, whether in XBRL, XML, or another structured data format, should be designed to eliminate ambiguity.

If the history disclosures were delivered in the format of a 50-column comma-separated values (CSV) file rather than as an XBRL file, casual users would be able to work with the data more easily. A CSV file can be readily imported into a spreadsheet or database without any programming knowledge, whereas XBRL files must be parsed programmatically before their contents can be analyzed. This “low technology” approach of CSV files remains useful when the structure of the data remains constant, and when both the sender and receiver have complete agreement on how to interpret the contents of that file.

The simplicity of CSV files has its benefits, but it also has its limits. In this particular use case, you can have many ratings associated with a single issuer and many obligors or updates to ratings. These one-to-many relationships can be represented in a CSV file, but at the cost of a tremendous growth in file size that may make it too large to load into a spreadsheet. Further, when there is a requirement to represent more richly described data (i.e., beyond that of a column label) or more complex disclosures and details, one should look to self-describing formats such as XBRL that are capable of satisfying these requirements.

---

152 Ibid.
154 Based on the author’s review of S&P, Moody’s and Fitch websites on May 14, 2016. At that time all three required a user id and password to access the rating history files.
156 One may consider a CSV file to be “self-describing data” if there is a header row and that row contains labels that identify the contents of their respective columns. It is more accurately understood as “self-identifying data” because it merely labels data by column rather than describing it, especially in comparison to the descriptive capabilities of XBRL.
OTHER OPEN DATA REPORTING MANDATES

Table 1 lists recent SEC final rules that required open data disclosures in XBRL or some other structured format, while Table 2 lists proposed rules that would impose an open data reporting requirement if finalized in their current form. Readers may notice the variety of data formats for these reports, including two not previously mentioned in this study: Financial products Markup Language (FpML) and Financial Information eXchange Markup Language (FIXML).

The use of multiple machine-readable data languages by the same agency is not itself problematic, provided the designers apply the advice given in the earlier “Recommendations for Improving the SEC’s XBRL Reporting Mandate for Financial Statements in Forms 10-K and 10-Q” section — specifically, designing an information model on underlying reporting or disclosure requirements rather than the printed documents.

The contents of these various reports are related to each other by virtue of their common reporting and disclosure requirements. Building an information model on these requirements enables the different specific technical implementations of taxonomies to work together, including the sharing of tags that logically should be held in common. Having a single, appropriate data format is ideal, of course, because it eliminates some technical obstacles; however, it is most critical to have a common information model based on the model of the actual subject of interest; in the cases pertaining to this study, the subjects of interest pertain to financial reporting and accounting standards.

Table 1: Other Recently Adopted SEC Open Data Filing Mandates

<table>
<thead>
<tr>
<th>Name of SEC Release</th>
<th>Date Adopted</th>
<th>Effective Date</th>
<th>Form</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Security-Based Swap Data Repository Registration, Duties, and Core Principles</td>
<td>2/11/2015</td>
<td>5/18/2015</td>
<td>SDR</td>
<td>XML</td>
</tr>
<tr>
<td>Money Market Fund Reform; Amendments to Form PF</td>
<td>7/23/2014</td>
<td>10/14/2014</td>
<td>N-MFP</td>
<td>XML</td>
</tr>
<tr>
<td>Crowdfunding</td>
<td>10/31/2015</td>
<td>5/16/2016</td>
<td>C</td>
<td>XML</td>
</tr>
</tbody>
</table>
These new mandates appear to be relatively small in scope, but a bill proposed in the 115th Congress would have greatly extended structured filing requirements. The Financial Transparency Act of 2017 (H.R. 1530), offered by Rep. Darrell Issa and 35 co-sponsors, would require several federal financial regulators to transform all of their reporting requirements to open data, though it does not explicitly specify any particular format.\(^\text{157}\)

### Table 2: Proposed SEC Structured Data Filing Mandates

<table>
<thead>
<tr>
<th>Name of SEC Release</th>
<th>Date Proposed</th>
<th>Form</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disclosure of Payments by Resource Extraction Issuers</td>
<td>12/11/15</td>
<td>SD</td>
<td>XBRL</td>
</tr>
<tr>
<td>Establishing the Form and Manner with which Security-Based Swap Data Repositories Must Make Security-Based Swap Data Available to the Commission</td>
<td>12/11/15</td>
<td>FpML, FIXML</td>
<td></td>
</tr>
<tr>
<td>Listing Standards for Recovery of Erroneously Awarded Compensation</td>
<td>7/1/15</td>
<td>N-CSR</td>
<td>XBRL</td>
</tr>
<tr>
<td>Investment Company Reporting Modernization</td>
<td>5/20/15</td>
<td>New Form</td>
<td>XML</td>
</tr>
<tr>
<td>Pay Versus Performance</td>
<td>4/29/15</td>
<td></td>
<td>XBRL</td>
</tr>
</tbody>
</table>

Source: Mike Willis, SEC Office of Structured Disclosure

These new mandates appear to be relatively small in scope, but a bill proposed in the 115th Congress would have greatly extended structured filing requirements. The Financial Transparency Act of 2017 (H.R. 1530), offered by Rep. Darrell Issa and 35 co-sponsors, would require several federal financial regulators to transform all of their reporting requirements to open data, though it does not explicitly specify any particular format.\(^\text{157}\)

The affected regulatory agencies include:

- Bureau of Consumer Financial Protection
- Commodity Futures Trading Commission
- Federal Deposit Insurance Corporation
- Federal Housing Finance Agency
- Federal Reserve System
- Municipal Securities Rulemaking Board
- National Credit Union Administration
- Office of the Comptroller of the Currency
- Securities and Exchange Commission

CONCLUSION

In an age of “big data” and “smart data” (and an age in which even Internet-aware kitchen appliances generate structured data), transitioning financial regulatory filers to database-friendly reporting formats should be seen as a common-sense reform. The needs of regulators, investors, and markets counsel extending the use of open data for financial reporting, not restricting or abolishing its use. Therefore, the question is when and how to proceed while minimizing costs and maximizing timely benefits. The SEC’s implementation of XBRL for the financial statements in Forms 10-K and 10-Q thus far has shown that a beneficial balance between costs and benefits is not so easily achieved.

Those who have worked in the information technology field understand the importance of gathering user requirements, the necessity of iterating through progressively improved versions of software and data models, and the cycle of building, testing, and revising. Financial regulators like the SEC, who may have a legal or financial analysis background, may be less familiar with these technical necessities, and less appreciative of the necessity for disclosure modernization through information technology. The early results in terms of excessive burden on filers and limited benefits to users are unsurprising given this circumstance; there was work to do, and this work would take attention and iterations over years before the purpose was achieved with ongoing operational efficiency.

Five years after US public companies began submitting their XBRL filings, conditions appear to be improving, especially as IT-literate professionals and accounting experts shape the on-going process. More filers leverage tools that integrate XBRL into the statement production process rather than treating it as an after-thought. Industry participants are automating more data quality tests, and the SEC is starting to use them to identify errors and to notify filers of their corrections. The FASB has reduced the number of elements to the US GAAP Financial Reporting Taxonomy, while the Data Quality Committee is providing guidance and validation rules that are freely available to prevent or detect inconsistencies and errors in the XBRL data. Market data aggregators like Bloomberg and Thomson Reuters are starting to use XBRL in their workflow processes, and hundreds of thousands of financial professionals have access to XBRL-derived data as a result.

These examples demonstrate how successful uses of XBRL transform regulatory reports into open data, enhancing the efficiency and effectiveness of regulatory oversight while decreasing the burden on regulated entities over the long term.

Rather than rollback its financial statement XBRL mandate, the SEC should address legitimate criticisms of its implementation. As recommended above, the SEC should:

- align its taxonomy with disclosure requirements, rather than trying to match the exact text in document-based financial statements;
- automate data quality tests;
- phase in iXBRL and eliminate unstructured text and traditional XBRL; and
- design this taxonomy, and future ones, for reuse.


These changes can be made relatively quickly, within the next few years, because they address policy problems, not technology ones.

Improvements must also be made to the SEC’s other XBRL reporting mandates, particularly the credit rating and mutual fund risk/return implementations. Credit rating XBRL disclosures should be placed in a central XBRL repository, lag times should be reduced, and the files should be made easier for researchers to consume. For the risk/return XBRL disclosures to become effective, filing times should be moved up and legacy formats phased out.

Going forward, new open data reporting requirements need better planning — ideally with the involvement of IT project management professionals, financial reporting experts, and data consumers — to minimize inconvenience for filers and maximize benefits both for the regulatory agency and for third-party users of the data.

There are other models to follow in both the US and in other countries. These examples demonstrate how successful uses of XBRL transform regulatory reports into open data, enhancing the efficiency and effectiveness of regulatory oversight while decreasing the burden on regulated entities over the long term.

Marc Joffe is a public policy analyst focusing on financial regulation and fiscal sustainability. He currently works as a Senior Policy Analyst at the Reason Foundation. His research has been published by the Mercatus Center at George Mason University, the Haas Institute for a Fair and Inclusive Society at UC Berkeley, the California State Treasurer’s Office and the California Policy Center. His op-eds have appeared in Bloomberg View, The Fiscal Times, The Hill, RealClearMarkets and The Guardian. Before becoming a policy analyst, Marc worked at number of financial institutions including JP Morgan Chase, CIBC and Moody’s Analytics.