

Speech

The Lessons of Structured Data



Commissioner Caroline A. Crenshaw

Washington D.C.

Nov. 10, 2021

Introduction

Thank you Mike [Schlanger] for that introduction, and thank you for inviting me to speak at this year's conference. I want to note at the outset that I appreciate all the work this organization does to support structured data, including enhancing the usefulness of the data in SEC filings. The work you do benefits the users of that data, including investors, academics, and of course, the SEC staff and other regulators. You have been vocal and energetic advocates for XBRL, and I appreciate that. So thank you.

And before I begin my remarks, I need to mention that the views I express today are my own and do not necessarily reflect the views of the Commission or its staff.

In putting together my remarks for today, I reflected on the history of XBRL and the SEC, including the series of decisions that the SEC made to require the use of structured data formats, and XBRL in particular, in data filed with us.

Our history with XBRL began when the Commission established a voluntary XBRL filing program for corporate financial statements in 2005. Then, in 2007, the voluntary program was expanded to permit mutual funds to submit their risk/return summary information as XBRL exhibits. These voluntary programs for operating companies and mutual funds were made mandatory in 2009.

The SEC followed with rules in 2009 requiring rating agencies to provide certain credit rating histories in XBRL on their websites. In 2018, the Commission adopted rules requiring operating company financial information and mutual fund risk/return summary information to be submitted in Inline XBRL, a specification of XBRL that is both human-readable and machine-readable. And in 2020, the Commission adopted rules that added Inline XBRL requirements for certain disclosures submitted by registered variable annuity and life insurance separate accounts, registered closed-end funds, and business development companies.^[1]

Requiring entities to change their filing and disclosure practices is not costless, and at every point, we carefully considered the costs and benefits of imposing these requirements. Today, looking back at this history, I believe it is clear that implementing structured data requirements has been a success. While there are certainly areas where data quality could be improved, XBRL has made it easier and less costly to extract, filter, compare, and analyze the information in SEC filings. XBRL facilitates the comparison of a company's information across time periods, against other companies, and between data in SEC filings and other agency filings. It allows for faster and more sophisticated analysis by regulators, investors, and academics. This increased usability has benefits for investors of all types.

I gave a speech earlier this year pointing out some areas in which I believe there are gaps in the information available to the SEC and investors.^[2] I believe one of the lessons of XBRL is that the Commission should not

hesitate to take action to fill those gaps, in order to get data we need to fulfill our mission of protecting investors, maintaining fair, orderly, and efficient markets, and facilitating capital formation.

Now, turning to the theme of today's event, I want to talk a bit about what XBRL delivers for the SEC and investors, how we can all work together to ensure shared access to higher quality data, and what XBRL might be able to deliver in the future.

What Does XBRL Data Deliver?

The Commission's implementation of XBRL requirements has allowed EDGAR to provide machine-readable data that have improved transparency in a number of ways. For example, XBRL has enabled automatic processing and analysis by software tools, which lowers costs and offers more timely insights. Users can access better and more granular information about these data, like the accounting codifications and guidance associated with it. Machine-readable languages like XBRL and iXBRL allow machine learning and artificial intelligence programs to leverage both numeric and narrative disclosures.^[3] It allows the automation of all manner of disclosure analysis – identifying what is and is not reported, identifying data quality errors, comparing results across data sets, performing other analytics, generating time series charting and benchmarking, and much more.

While this is all fairly technical, for me anyway, the bottom line is that it makes these data more useful. And we know that XBRL data are used by investors – institutional investors of course, but also retail investors, who rely on tools and analyses that are facilitated by structured data.^[4] The SEC is, of course, a user of the data, and I know that later today you will hear more about some of our recent use cases. But we also know that XBRL data are used by other regulatory agencies, including the IRS, Treasury, and the Census Bureau, to name a few.^[5] XBRL data are used in academic research,^[6] and are also used by financial analysts,^[7] news media,^[8] data aggregators,^[9] operating companies,^[10] and a host of others.

All of this user activity adds up to more market transparency and more efficient markets. For example, since the implementation of the XBRL mandate, we have seen stock prices become more reflective of firm-specific disclosures;^[11] we've seen increased quantitative disclosure from firms;^[12] and we've seen decreased earnings smoothing.^[13] It also adds up to fairer, more competitive markets. Research indicates that XBRL disclosures reduce the advantages enjoyed by insiders, relative to non-insiders,^[14] as well as the advantages of locals relative to non-locals.^[15]

Some research has indicated that the use of XBRL leads to more equal outcomes between large investors and analysts and small ones,^[16] and reduces the advantages enjoyed by institutional investors as compared to individuals.^[17] Research has also suggested that XBRL leads to a better-informed investing public, as a result of improved financial analysis.^[18] We know that this is especially relevant to retail investors, who often rely on analysts and media to inform them about the markets.^[19]

XBRL disclosures may also facilitate capital formation, as some academics have found that companies that use it enjoy a lower cost of capital.^[20] This is particularly true for smaller companies, which tend to receive more analyst attention following XBRL adoption.^[21] It also results in higher investment efficiency for companies,^[22] and enables improved performance benchmarking and acquisition analysis.^[23]

Finally, the SEC's use of the data facilitates better investor protection; our staff leverages structured data tools in enforcement, examinations, and policymaking.^[24] And I believe you will be hearing more about that later today.

I know that was a bit of a laundry list, and thank you for bearing with me. Ultimately, the point I want to make is that, while there are costs associated with complying with the XBRL mandate, the benefits are well-documented and extensive. And that is why I feel very comfortable saying that the story of the XBRL mandate is a successful one.

How Can We Deliver Better Data?

However, while I believe XBRL data are delivering myriad benefits, there is room for improvement in terms of the quality and accuracy of the data. Some users have found material error rates in data tagged in our filings, including errors in tags that are likely to be crucially important to investors like Revenues, Net Income, and Assets, and scaling errors that can be impactful.^[25]

Both filers and the SEC have roles to play in mitigating these errors. The primary responsibility lies with filers, of course, and there are many tools available to aid with the submission of high-quality, accurate data. For example, EDGAR provides validation warnings which flag data quality issues, such as the use of outdated tags.[\[26\]](#) When submitting filings to the Commission, filers should ensure that they address those warnings.

In addition, the organization hosting us today, XBRL US, provides data quality validation rules registrants can use before submitting filings.[\[27\]](#) Use of these is not required but can assist in identifying errors, and I would encourage filers to take advantage of this free resource. XBRL US also publishes information about which filings have data quality errors, as measured against their validation rules – that information is available to the public, and I would encourage filers to take a look.[\[28\]](#)

XBRL allows for the use of individualized, custom tags. Custom tags have informational value, when used appropriately, and can increase investor understanding. However, some filers may overuse these tags. Filers should make an effort to use standard tags, and only use custom tags when appropriate – that is, when no standard tag is applicable.[\[29\]](#)

The SEC's Office of Structured Disclosure publishes staff guidance and regular data quality reminders, that I encourage filers to review.[\[30\]](#) Recent alerts have flagged issues like scaling errors on public float data and incorrect period end dates. And of course, if filers or others have technical questions on structured data and data quality, they can always reach out to the Office of Structured Disclosure by email.[\[31\]](#)

As I noted earlier, data quality is primarily the responsibility of filers. However, there are things that the SEC can do to help, as well. In addition to the resources offered by the Office of Structured Disclosure that I just mentioned, the SEC staff has released public comment letters regarding deficiencies in XBRL filings. The SEC staff should utilize this tool as much as possible, to help highlight and address common errors. The SEC could also consider expanding the requirements for auditor assurance, to provide for more third-party verification and validation of tags.

Delivering More

Looking ahead, I want to briefly mention a few areas where I think XBRL can play a role in delivering even greater benefits to investors and the market.

In her remarks at this conference last year, my fellow Commissioner Allison Herren Lee noted that structured data could play a role in making disclosures of climate change and other ESG risks and impacts usable and comparable.[\[32\]](#)

Since then, we have requested comment on climate change disclosures in particular, and have heard from a number of commenters who support structuring that data.[\[33\]](#) I look forward to working with the staff in carefully considering those comments and the potential benefits.

Similarly, Commissioner Lee mentioned the potential benefits of using structured data in Form N-PX, which provides information about proxy voting by investment funds. Since then, we have re-proposed amendments to that form, including a requirement to report information in a structured data language.[\[34\]](#) Again, commenters have supported that potential requirement, and I look forward to working with the staff to consider potential approaches.[\[35\]](#)

Now, as I've discussed at length today, the potential benefits of tagging data are extensive. So we at the SEC should continue to investigate where else data structuring can improve our disclosure ecosystem. The tagging of narrative disclosures, even just block tagging,[\[36\]](#) could enable data users to more easily extract and compare non-structured disclosures, like management discussion and analysis, earnings reports, and executive compensation. This could be relevant in the context of ESG disclosures, SPAC disclosures, and elsewhere.

Finally, I would be remiss not to mention the potential benefits of incorporating the Legal Entity Identifier (LEI) into more of our forms and filings. As most of you undoubtedly know, the LEI is a code that provides a single, unique, international identifier for legal entities.[\[37\]](#) As such, it facilitates the reliable, consistent identification of entities within and across data sets.

Last year, the LEI in XBRL Working Group published an LEI XBRL taxonomy enabling its use in XBRL applications to unambiguously identify companies.^[38] The inclusion of LEIs in XBRL data has the potential to increase the usefulness of these data in SEC filings in a number of ways – for example, consistently identifying relevant entities in supply chains, or linking information on an entity across multiple regulatory data sets.

While the SEC has taken steps to incorporate LEIs into our filings,^[39] I believe we should continue to leverage their benefits by incorporating them into our forms and filings wherever it makes sense to do so. The ability to use LEIs in XBRL data only increases their potential utility for users of our data.

Conclusion

Before I conclude my remarks, I want to return to what I said earlier about reflecting on the lessons of XBRL, after nearly 15 years. As I've outlined today, I believe that, while there is always room for improvements, the story of XBRL and the SEC is a successful one. The Commission's decision to require the use of structured data in our filings has had a host of benefits for investors and the markets.

In the speech I gave earlier this year, I pointed out some areas in which I believe there are gaps in the information available to the SEC and investors.^[40] And, I believe that a key lesson of the SEC's history with XBRL is that the Commission should not hesitate to act to ensure that we have the information we need to fulfill our mission.

The areas I identified for attention included private markets, trade and order data – specifically, the urgency of completing the Consolidated Audit Trail, or CAT – and investor testing of certain disclosures. In each of those areas, I think our experience with XBRL offers valuable lessons.

For example, with respect to private markets, I believe it should be a priority to act to finalize the changes the SEC proposed in 2013 to strengthen filing and disclosure requirements. This would provide some needed visibility into private issuers and offerings.

We should also take action to get the information we need about the effectiveness of Form CRS and the disclosures required under Regulation Best Interest, by engaging in investor testing of the actual forms and disclosures that investors receive in order to determine whether or not they are effective.

Finally, with respect to the CAT, I think it is vital to ensure that it reaches its full potential as a tool for understanding and analyzing the markets we regulate. The CAT operates pursuant to a National Market System plan, under which the national securities exchanges and FINRA share responsibility for its operation and administration. If that approach is not working, we should not hesitate to take whatever action may be necessary to accomplish the objective of a complete, accurate, and accessible source of market data, including further rulemaking, if needed.

With XBRL, and with all reporting and disclosure requirements, we need to be cognizant of the impact on market participants. However, as I've noted in the past, the lack of useful data has a cost as well.^[41] In the case of XBRL and structured data formats more generally, the Commission has taken a number of actions over the years to ensure that investors, academics, the SEC staff and other regulators, and the public more generally can all benefit from the data filed with us.

While there is more work to be done, I believe our efforts to incorporate structured data formats and open-source identifiers into the data filed with us have been a success. I appreciate all the work of this organization in helping us get to where we are, and I look forward to working with you as we continue to use these technologies to improve transparency for investors and others.

Thank you so much, and I look forward to your questions.

[1] See Securities and Exchange Commission, [Structured Disclosure at the SEC: History and Rulemaking](#) (May 21, 2020).

[2] Caroline Crenshaw, [Mind the \(Data\) Gaps](#) (May 14, 2021).

[3] See, e.g., Baranes et al., [Earning Movement Prediction Using Machine Learning-support Vector Machines \(SVM\)](#) (2019), *Journal of Management Information and Decision Sciences*; Singh, [Blockchain and XBRL: The](#)

[Myth](#), CFA Institute (2020).

[4] See, e.g., Goldman Sachs Asset Management, [First Take: From Flat to Down](#), '19 Pension Review (2020); Cong et al., [Are XBRL Files Being Accessed? Evidence from the SEC EDGAR Log File Dataset](#), Journal of Information Systems (2018); Blankespoor, Elizabeth and deHaan, Ed and Marinovic, Ivan, [Disclosure Processing Costs, Investors' Information Choice, and Equity Market Outcomes: A Review](#), Journal of Accounting & Economics (JAE), Forthcoming (January 2020).

[5] See, e.g., Toppan Merrill, [100% XBRL Coverage Has Transformed SEC Review and Enforcement](#) (November 19, 2019); XBRL US, [FDIC Reporting](#).

[6] See, e.g., Hoitash et al., [Do Sell-Side Analysts' Qualifications Mitigate the Adverse Effects of Accounting Reporting Complexity?](#), SSRN (2019); Hoitash et al., [An Input-Based Measure of Financial Statement Comparability](#), SSRN (2018); Henselmann et al., [Content analysis of XBRL filings as an efficient supplement of bankruptcy prediction? Empirical evidence based on US GAAP annual reports](#), Working Papers in Accounting Valuation Auditing (2012).

[7] See, e.g., Morgan Stanley Research, [Who's Using XBRL Data and Why: Case Studies](#) (2017).

[8] See, e.g., Trentmann, [Companies Adjust Earnings for Covid-19 Costs, but Are They Still a One-Time Expense?](#) The Wall Street Journal (2020).

[9] See, e.g., XBRL US, [How Third Party Data Providers Use Structured Data and Why](#); XBRL US, [Financial Fundamentals Analysis – What Analysts Can Do with Structured Data](#) (March 2016).

[10] See, e.g., Berkman, [XBRL: What are the Benefits?](#), Financial Executives International (2019); Rao et al., [Using XBRL and big data to improve decision-making](#), Financial Management (2020).

[11] See, e.g., Huang et al., [Information Processing Costs and Stock Price Informativeness: Evidence from the XBRL Mandate](#), Australian Journal of Management (2020); Kim et al., [Information-Processing Costs and Breadth of Ownership](#), Contemp Account Res. (2019); Efendi et al., [Do XBRL filings enhance informational efficiency? Early evidence from post-earnings announcement drift](#), Journal of Business Research (2014),

[12] See Blankespoor, [The Impact of Information Processing Costs on Firm Disclosure Choice: Evidence from the XBRL Mandate](#). Journal of Accounting Research (2019).

[13] See, e.g., Kim et al., [Does XBRL Adoption Constrain Earnings Management? Early Evidence from Mandated U.S. Filers](#), Contemporary Accounting Research/Accepted Articles (2019).

[14] See Huang et al., [Insider Profitability and Public Information: Evidence From the XBRL Mandate](#), SSRN (2019). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3455105 ; Zhu, [The Effect of XBRL on Insider Trading Profitability](#), Erasmus Univeriteit Rotterdam (2018).

[15] See Li et al. (2020), [The Impact of XBRL Adoption on Local Bias: Evidence from Mandated U.S. Filers](#), Journal of Accounting and Public Policy (2020).

[16] See, e.g., Bhattacharya et al., [Leveling the Playing Field between Large and Small Institutions: Evidence from the SEC's XBRL Mandate](#), The Accounting Review (2018).

[17] See Zhu, [The Effect of XBRL on Insider Trading Profitability](#), Erasmus Univeriteit Rotterdam (2018).

[18] See, e.g., Liu et al., [XBRL's Impact on Analyst Forecast Behavior: An Empirical Study](#), Journal of Accounting and Public Policy (2014); Felo et al., [Can XBRL detailed tagging of footnotes improve financial analysts' information environment?](#), International Journal of Accounting Information Systems (2018).

[19] See, e.g., Kim et al., [Investor Sentiment, Stock Returns, and Analyst Recommendation Changes](#), SSRN (2019); Lawrence et al., [Investor Demand for Sell-Side Research](#), The Accounting Review (2016); Kothari et al., [The Effect of Disclosures by Management, Analysts, and Business Press on Cost of Capital, Return Volatility, and Analyst Forecasts: A Study Using Content Analysis](#), The Accounting Review (2009).

[20] See, e.g., Lai et al., [XBRL adoption and cost of debt](#), International Journal of Accounting & Information Management (2015); Ra et al., [XBRL Adoption, Information Asymmetry, Cost of Capital, and Reporting Lags](#), iBusiness (2018).

- [21] See, e.g., Li et al., [Does XBRL Adoption Reduce the Cost of Equity Capital?](#), SSRN (2012).
- [22] See, e.g., Feng et al. [Information processing costs and firms' investment efficiency: evidence from the SEC's XBRL mandate](#), SSRN (2020).
- [23] See, e.g., Berkman, [XBRL: What are the Benefits?](#), Financial Executives International (2019); Rao et al., [Using XBRL and big data to improve decision-making](#), Financial Management (2020).
- [24] See, e.g., Toppan Merrill, [100% XBRL Coverage Has Transformed SEC Review and Enforcement](#) (November 19, 2019).
- [25] See Calcbench, [The Quality of XBRL Filings](#) (2014) (“[T]here are a non-trivial number of errors in tags which are likely to be heavily used by analysts and investors (e.g., Revenues, Net Income, Assets, etc.)”); see also XBRL US, [Aggregated Real-time Filing Errors](#).
- [26] Securities and Exchange Commission, [EDGAR XBRL Validation Warnings](#) (June 2021).
- [27] XBRL US, [Approved Validation Rules](#) (October 2021).
- [28] XBRL US, [Filing Results and Quality Checks](#).
- [29] See 17 CFR 232.405(c)(1)(iii)(B) (“An electronic filer must create and use a new special element if and only if an appropriate tag does not exist in the standard list of tags for reasons other than or in addition to an inappropriate standard label.”) (emphasis added).
- [30] Securities Exchange Commission, [Staff Observations, Guidance, and Trends](#) (October 2021).
- [31] For technical questions on structured data and data quality, contact the Office of Structured Disclosure at Structureddata@sec.gov.
- [32] Allison Herren Lee, [The Promise of Structured Data: True Modernization of Disclosure Effectiveness](#) (November 17, 2020).
- [33] [SEC.gov | Comments on Climate Change Disclosures](#)
- [34] See [Enhanced Reporting of Proxy Votes by Registered Management Investment Companies; Reporting of Executive Compensation Votes by Institutional Investment Managers](#), Release Nos. 34-93169; IC-34389 (December 14, 2021).
- [35] See <https://www.sec.gov/comments/s7-11-21/s71121.htm>
- [36] See XBRL US, [Comment letter re: Management's Discussion and Analysis, Selected Financial Data, and Supplementary Financial Information](#), File Number S7-01-20 (April 28, 2020).
- [37] See Global Legal Entity Identifier Foundation (GLEIF), [Introducing the Legal Entity Identifier](#).
- [38] See XBRL, [LEI Taxonomy finalized](#) (July 3, 2020).
- [39] See, e.g., [Regulation SBSR – Reporting and Dissemination of Security-Based Swap Information](#), Release No. 34-74244 (Feb. 11, 2015), 80 FR 14439 (Mar. 19, 2015); [Investment Company Reporting Modernization](#), Release No. 33-10231 (Oct. 13, 2016) 81 FR 81870 (Nov. 18, 2016).
- [40] See Caroline Crenshaw, [Mind the \(Data\) Gaps](#) (May 14, 2021).
- [41] *Id.*