

November 15, 2021



1211 Avenue of the Americas
19th Floor
New York, NY 10036
Phone: (202) 448-1985
Fax: (866) 516-6923

Office of Shared Solutions and Performance Improvement (OSSPI)

To: Whom It May Concern:

RE: RFI on Behalf of the Federal Chief Data Officers Council, Docket No. 2021

We appreciate the opportunity to provide input to the Chief Data Officers (CDO) Council mission and focus. We support the Council's vision to improve government mission achievement and increase the benefits to the Nation through improvement in the management, use, protection, dissemination, and generation of data in government decision-making and operations.

I am the CEO of XBRL US, a nonprofit data standards organization. Our mission is to improve the efficiency and quality of reporting in the U.S. by promoting the adoption of business and government data standards. Better reporting leads to greater transparency and accountability, and improves the ability of government agencies to set useful policies. Better reporting promotes efficiency improvements that save money and time for those reporting, collecting, and using data.

XBRL US members include public companies, accounting firms, software, data, and service providers, as well as other nonprofits and standards organizations. XBRL US is a jurisdiction of XBRL International¹, the nonprofit consortium responsible for developing and maintaining the technical eXtensible Business Reporting Language (XBRL) specification, which is a free and open data standard widely used around the world for reporting by public and private companies, as well as government entities.

This letter recommends the adoption of data standards to support the Council's mission and goes on to address sections in the RFI.

Adopt a Single Data Standard Across Federal Agencies

We urge the Council to leverage open, nonproprietary data standards to maximize the value of collected data and reduce the cost of data reporting, collection, and use. The standard that we support, eXtensible Business Reporting Language (XBRL) has been adopted by eighty regulators around the world and is used today in 184 government data collection programs². Federal agencies in the U.S. that require reporting in XBRL include the Securities and Exchange Commission (SEC), the Federal Deposit Insurance Corporation (FDIC) and the Federal Energy Regulatory Commission (FERC). Effective regulatory data collection programs focus on results by ensuring that the standards adopted are:

¹ XBRL International: <https://xbrl.org>

² XBRL Project Directory: <https://www.xbrl.org/the-standard/why/xbrl-project-directory/>

- | | |
|---|---|
| ✓ Open, nonproprietary (free) | ✓ In widespread use (otherwise, it's not a standard) |
| ✓ The right fit for the data reported | ✓ Long-term solution that adapts easily (inexpensively) to change |
| ✓ Adaptable to open source and commercial tools | ✓ Relies on a single data model documented in a machine-readable taxonomy |

When government regulators implement effective data standards programs³, they get results:

- Machine-readable, automatable, timely data
- High quality data, further checked by universal data integrity checks
- Ability to commingle multiple data sets
- Reduced cost for reporting entities, data collectors, and users
- Reduced reporting burden (eliminates duplication, increases data checking)

Unfortunately, we often see well-meaning regulators adopt short-cuts which may make the initial implementation process easier and limit change for reporting entities and data collectors, but do not yield desired results. In fact, these approaches are substantially more costly in the long run:

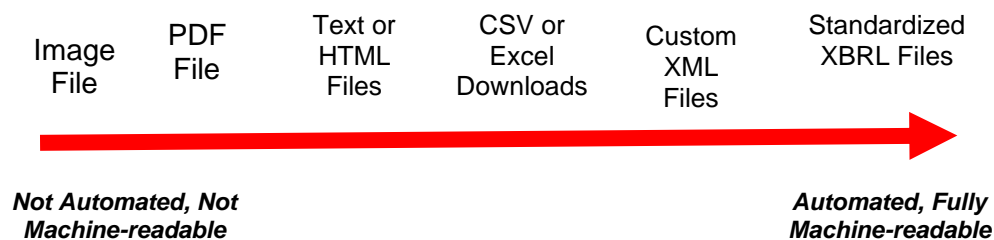
- Collecting data in spreadsheets. Cannot be automated. Requires manual review, cutting & pasting. Results in poor data quality, lack of timeliness.
- Developing a single system or product. Locks regulator into single vendor. Expensive to change reporting requirements.
- Building a custom schema. Requires market to build custom tools to create, collect, and use data. Expensive to change reporting requirements.

Recommendations on Section 3: Data Inventories

How do you find Federal data?

Federal data is published in different formats, with various levels of accessibility, as shown on the “automation spectrum” figure below. Data in an image file or a PDF file is not machine-readable, and therefore consumption of data embedded in these files cannot be automated. The Federal Audit Clearinghouse today collects tens of thousands of Single Audit Report PDF files annually, that are up to several hundreds of pages in length. These filings are resistant to standardized data extraction and typically require human review. Similarly, certain portions of the Single Audit Report are made available in large tab-delimited text files. While the text files provide a slightly higher level of automation, each file must be extracted from zip archives, and the data manually interpreted by reviewing rows and columns before analysis can begin.

³ Infographic “Effective data standards programs focus on results”: <https://xbrl.us/infographic/data-standards-regulators/>



CSV or Excel downloads are better than text files in terms of automation, yet still require manual review before use. The Environmental Protection Agency (EPA) makes facilities emissions data available in downloadable spreadsheets. This is highly valuable, complex data but to use it requires cross-review of other spreadsheets containing parent company data, and again, manual interpretation of row and column headers to understand the data reported.

Custom XML files, while they are machine-readable, require the building of custom applications to report, and extract and analyze data. These custom applications can only be used for that single custom data set.

Standardized XBRL files, however, are unambiguously machine-readable; data transported in XBRL format can be read direct from machine-to-machine with no need for manual interpretation or extraction. Because XBRL is a widely used standard, tools that work with one implementation will work with any other XBRL program, keeping the cost of analysis low and allowing for the commingling of data. As mentioned earlier, XBRL formatting is required for financial data reported by public companies to the SEC, banks to the FDIC, and utilities that report to the FERC. Tools that work with SEC data, also work with FERC and FDIC data. A bigger market opportunity encourages competition.

Because a lot of federal data today is published in different formats, publishing agencies often maintain their own data silos, which limits the ability to commingle information and access richer data sets.

How can data inventories best support how you identify Federal data that is valuable for your own use case? How could existing platforms better support access to Federal data?

Data inventories should be searchable in a standardized way, and allow for searching across inventories. For example, a user may wish to identify all the instances of reporting by a particular company across agencies; or may search all the data stores that contain information about tax obligations of companies.

Data inventories must provide sufficient explanatory metadata to enable highly granular searching, to allow users to search across all agencies, entities, and data types automatically, without manual searching required.

What is the best implementation of a data inventory you have seen? What are the characteristics that made it so successful?

The SEC program collects and posts public company financials in XBRL format every quarter. They make RSS feeds available for anyone who wishes to extract the freely available data. Because of the consistent nature of this data, data aggregators extract and store the data as soon as it is posted by the SEC; and serve the data to users within minutes of its receipt by the SEC. My organization, with a staff of four, maintains our own database of SEC financial data with continuous updates and ease of access because of the structured nature of the content.

One important benefit to the SEC is that they simply post the machine-readable files. There is no need for further manipulation or extraction or querying tools to make the data accessible to the public. Because the data is so easy to use, the market takes care of that for the regulator.

To date, inventories have relied on manual work to generate and maintain metadata. What best practices and tools are available to automate and reduce the manual workload associated with inventories?

A taxonomy is a digital dictionary of terms that serves as the single data model for what needs to be reported. Embedded in the taxonomy is metadata associated with each reported concept. For example, Revenues would have a monetary data type, a duration period type, and an associated definition. The taxonomy is referenced by applications to prepare reports, collect data, and extract and analyze data. It can also be used as the source of the data inventory because the taxonomy describes what is allowed to be reported.

Recommendations on Section 4: Data sharing

What best practices could statistical agencies and non-statistical agencies use to better partner?

Sharing of data across agencies is critical to the efficient functioning of government and setting of policies. Data management should allow automated access to cross-agency, cross-entity data with appropriate security safeguards.

Data sharing is enabled when the data structure is the same across reporting entities. Today for example, public utilities, bank institutions and public companies all report in XBRL format. Very different data reported following different accounting guidelines, but because the underlying structure of the data is the same, the same tool used to extract SEC data, can also be used for FDIC and FERC data. Reporting in a single structure means that commercial and open-source tools can all compete to prepare, collect, extract, and analyze data, thus lowering costs for all involved. Maintaining the same structure for the data allows for consistent linking across agencies.

Data sharing is also facilitated by simplifying the use of identifiers. For example, our organization, XBRL US, is a small nonprofit, and yet we have seven separate identifiers, each for a different purpose:

LEI - [9845008D4851D6A7M390](#)
EIN - [20-5592157](#)
UEI - NMBAWP3BM235
CAGE - 55WS6

DUNS - 801992293
CIK - 00892090141
CID – In Process

The ability to track an entity and to understand their relationship to other entities is critical to evaluate investment and business risk. That is why we strongly recommend that the CDO Council require the adoption of the Legal Entity Identifier (LEI) for all reporting entities. The LEI is becoming much more commonly used worldwide and we are seeing increasing use in the United States as well. Unlike DUNS numbers, LEIs are freely redistributable.

How can the Federal government engage with private sector data providers in a way that maximizes the ability to use the data or data derivatives across multiple agencies? How might we achieve this while ensuring a viable business model for data providers?

Data aggregators and analytics providers prefer access to structured, machine-readable data because it is significantly easier (and less expensive) to process. As noted by Morningstar in a recent video⁴, *“Extracting data from an HTML document takes at least 20 minutes, from a good quality PDF, takes around 30 minutes, from an image around 50 minutes. Data pulled from an XBRL file though, can be extracted in 1 to 2 seconds... lets us focus on better analytics rather than scraping data from documents.”*

Lower processing costs for data providers translates into lower costs for investors, analysts, government agencies, and other users of data. The more data available in structured, XBRL format, the less expensive it becomes for all parties involved.

Recommendations on Section 6: Ethics and Equity

How can we leverage Federal Data Ethics to improve trust and transparency?

Timely, high integrity data is critical to assisting ethics and equity in decision-making. For federal agencies to meet these standards, they need access to high quality, timely (not years out of date) information to make the right decisions. Good quality data provides transparency and accountability to the public.

Recommendations on Section 7: Technology

What frameworks should agencies use to evaluate their existing data infrastructure and to modernize technology with capabilities that break down organizational data silos and ensure the best available data is available?

An agency-wide structured data standards framework is the best approach to break down data silos, encourage interoperability, and ensure the availability of high quality, consistent data sets that can be shared across agencies and made available to the public.

⁴ XBRL for Analysts and Investors: <https://xbrl.us/news/analyst-video/>

A good quality data standards program provides the needed framework to leverage other technologies, like machine-learning and Artificial Intelligence platforms. Structured data is a needed ingredient to drive useful outcomes from these technologies. The appropriate data standard also allows for the linking of data through graphs.

How do we get there?

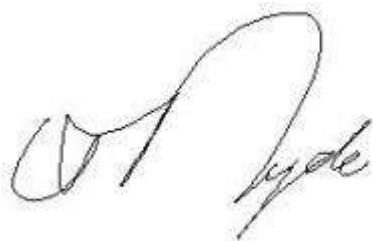
We are already there.

Structured data standards have resolved the issues of data sharing, inventories, and maintaining equity and maturity, in hundreds of programs around the world. The proof is there. To put these successful programs in place requires thought leaders who have the vision to enact change that leads to a long-term solution. The CDO Council can be that change agent. We urge members of the Council to consider the many effective regulatory programs around the world including these in the U.S.:

- 6,000 public companies and numerous investment management companies reporting quarterly and annually to the SEC
- 5,000 banks reporting to the FDIC
- Six hundred public utilities reporting to the FERC

We are interested in scheduling a follow up call with the various committees of the Council to address these issues, and to respond to any questions you may have. Contact me by emailing campbell.pryde@xbrl.us or at (917) 582-6159.

Respectfully,

A handwritten signature in black ink, appearing to read 'Campbell Pryde', is positioned above the typed name and title.

Campbell Pryde,
President and CEO