From:	Anna Kwok
To:	Comments
Cc:	David Haila; Brandon Ziegler
Subject:	[redacted] Comments on RIN 3064-AF96
Date:	Friday, October 18, 2024 6:28:21 PM
Attachments:	Workiva FDTA Joint Data Standards Comment Letter.docx.pdf

[redacted]

Please find attached our comments in PDF format on the Financial Data Transparency Act Joint Data Standards, RIN 3064-AF96, submitted on behalf of David Haila, EVP, Chief Technology Officer of Workiva. Thank you.

Anna Kwok

Anna Kwok
Structured Data Initiatives
Workiva Inc.

[redacted]

Via electronic submission

October 18, 2024

Chief Counsel's Office, Attention: Comment Processing, Office of the Comptroller of the Currency, 400 7th Street SW, Suite 3E-218, Washington, DC 20219

Ann E. Misback, Secretary, Board of Governors of the Federal Reserve System, 20th Street and Constitution Avenue NW, Washington, DC 20551

James P. Sheesley, Assistant Executive Secretary, Attention: Comments/Legal OES (RIN 3064-AF96), Federal Deposit Insurance Corporation, 550 17th Street NW, Washington, DC 20429

Melane Conyers-Ausbrooks, Secretary of the Board, National Credit Union Administration, 1775 Duke Street, Alexandria, Virginia 22314-3428

Comment Intake—FDTA-INTERAGENCY RULE, c/o Legal Division Docket Manager, Consumer Financial Protection Bureau, 1700 G Street NW, Washington, DC 20552

Clinton Jones, General Counsel, Attention: Comments/RIN 2590-AB38, Federal Housing Finance Agency, 400 Seventh Street SW, Washington, DC 20219

Christopher Kirkpatrick, Secretary of the Commission, Commodity Futures Trading Commission, Three Lafayette Centre, 1155 21st Street NW, Washington, DC 20581

Vanessa A. Countryman, Secretary, Securities and Exchange Commission, 100 F Street NE, Washington, DC 20549-1090

Chief Counsel's Office, Attention: Comment Processing, Office of Financial Research, Department of the Treasury, 717 14th Street NW, Washington, DC 20220

RE: Financial Data Transparency Act, Docket ID OCC-2024-0012
Docket No. R-1837 and RIN 7100-AG-79
RIN 3064-AF96
Docket Number NCUA-2023-0019, RIN 3133-AF57
Docket No. CFPB-2024-0034, RIN 3170-AB20
RIN 2590-AB38
Financial Data Transparency Act Joint Data Standards Rulemaking, RIN number 3038-AF43
File Number S7-2024-05
RIN [1505-AC86]

Dear Financial Data Transparency Act Implementing Agencies:

Workiva appreciates the opportunity to provide comments in response to the proposed Financial Data Transparency Act (FDTA) Joint Data Standards. We support establishing properties-based joint data standards for the agencies' data collections, and encourage the agencies to prioritize data standards and formats based on the established, machine-readable XBRL standard.

Founded in 2008, Workiva is a cloud-computing software provider and filing agent serving public registrants, financial services firms, and governments across their financial, non-financial, and risks and compliance processes, data, and reporting. Workiva provides reporting solutions that support structured data requirements for the SEC, FERC, ESMA, as well as statutory authorities in the UK, Ireland, Netherlands, Germany, South Africa. We serve over 2,800 of 8,400 operating companies reporting to the SEC, based on our dataset of unique CIKs that filed a 10-K, 20-F, or 40-F in 2023, and we have extensive experience in implementing data standards across jurisdictions. We understand the fundamental needs for transparency, interoperability, and auditability of data for our customers, and we appreciate the opportunity to provide input based on our global experience and structured data expertise.

The main objectives of FDTA resonate with our technology approach supporting regulated entities for both internal and external reporting.

> *"To amend securities and banking laws to make the information reported to financial regulatory agencies electronically searchable,* ***to further enable the development of regulatory technologies and artificial intelligence applications****, to put the United States on a path towards building a comprehensive Standard Business Reporting program* ***to ultimately harmonize and reduce the private sector's regulatory compliance burden****, while enhancing transparency and accountability, and for other purposes."* (S.4295[1])

Most Workiva customers have multiple reporting obligations. As regulated entities, they welcome data interoperability because they benefit when the same data standards work across multiple regulators, standardizing the way the reporting obligations are fulfilled in order to streamline processes and increase efficiency. The more the same data standards can be reused, the more benefits for a standardized reporting experience while producing interoperable and high quality data.

We are pleased to provide further input on the joint proposed rule with a focus on the following main areas, each of which is detailed in subsequent portions of this comment letter:

1. **Schema/taxonomy standards:** We recommend the agencies prioritize the establishment of the high-resolution semantic standard of XBRL as the machine-readable syntax for the schema/taxonomy standards based on its native structure to communicate metadata and meaning consistently, making it

[1] https://www.congress.gov/bill/117th-congress/senate-bill/4295/text

an efficient standard for producing interoperable data. Other formats such as XML, on the other hand, are typically customized for each information collection, leading to efficiency and interoperability loss in implementing the data standard.

2. **Data transmission standards**: We recommend the agencies prioritize the establishment of an XBRL-compatible syntax for the data transmission standards, that is, XBRL-XML, XBRL-JSON, XBRL-CSV, and Inline XBRL. This helps retain the rich data model of XBRL and the immediate utility of its open source libraries for automated data transmission and extraction, and enables flexible transmission formats based on individual data collection needs. In contrast, automatic data extraction from PDF/A requires customized coding for each use case, and is not as proven for high data load in complex business reporting.
3. **Accounting and reporting taxonomies standards**: We recommend the agencies focus on the establishment of domain/topic-specific taxonomies, for example, the US GAAP taxonomy as the established taxonomy standard for GAAP-based reporting. Domain/topic-specific schemas/taxonomies promote reuse, increasing consistency and data interoperability while reducing siloed or overlapping data standards.

We further expand the details of our input on the agencies' invitations for comments below.

1. The Agencies invite comment on the incorporation of the PRA definition of "collection of information" for purposes of the proposed rule.

We support the incorporation of the PRA definition of "collection of information" for purposes of the proposed rule. XBRL and Inline XBRL formats are for reports that are based on widely-used industry data standards, establishing a foundation for machine-readable reports. Expanding the movement from documents to machine-readable data should be a broad effort by the agencies. We believe that the scope of the PRA already aligns with the efforts of our users to move more of their reports to revolve around data, thus the scope of the PRA matches industry trends. Additionally, though not explicitly stated in the proposed rule, we also interpret the inclusion of the PRA as recognizing that the data standards will be for data points in a report, mirroring the current approach for machine-readable compliance reporting rules such as for the 10-Q and the 10-K at the SEC. Our users collect and work with individual classifications of data (i.e., a data point), thus establishing data standards as the FDTA is meant to do will necessarily be at a similar level of detail.

2. The Agencies invite comment on the establishment of the LEI as the legal entity identifier data standard in the proposed joint rule and on other options for the legal entity identifier data standard.

We support establishment of the LEI as the legal entity identifier data standard. The Legal Entity Identifier (LEI) is already in use in reporting today with an XBRL taxonomy in widespread use which Workiva already supports in its solutions. For us and our users there is essentially no incremental burden to widen the use of the LEI for machine readable data via the LEI taxonomy.

3. The Agencies also request comment on the use of the LEI to identify legal entities related to the filer of a particular report, such as a subsidiary or parent of the filer.

Likewise, as our solutions already provide for LEI support in XBRL, for us and our users there is essentially no incremental burden to widen the use of the LEI taxonomy for information related to subsidiaries and parents of filers. Anecdotally, many of our users are entities within a network of ownership. Tasks such as multi-entity consolidation, a common use case among our users, already require a means of entity identification. Further, many of our users have operations outside the US in countries that already require the use of the LEI, and so these users utilize the LEI to represent their network of entity ownership in a manner that is consistent both in terms of definition and in terms of technology support for the standard.

4. The Agencies invite comment on the establishment of other common identifiers in the proposed rule.

We support the adoption of the ISO standards in the proposed rule. The ISO standards for date (ISO 8601), country (ISO 3166 and GENC), and currency (ISO 4217) are already incorporated into XBRL standards today for structured data reporting, which Workiva also already supports. We would also plan to support any additional standards for identifiers in our platform in a timely manner so that our users may also comply.

In addition, where non-government organization (NGO) taxonomies exist for a domain/topic-specific standard, for example, the XBRL International's Country and Currency taxonomies, we would encourage migration to those NGO taxonomies. Avoiding agency-specific taxonomies where common domain/topic-specific taxonomies already are available could help promote data interoperability not only across the agencies but also across other jurisdictions globally where the regulated entities may have dual reporting obligations. For example, an IFRS filer submitting to the SEC and also the European Securities and Markets Authority (ESMA).

Moreover, reuse of standards increases consistency and data interoperability. The more joint standards like these are adopted, the lower the compliance burden on the regulated entities to satisfy all reporting obligations.

5. The Agencies invite comment on the proposed establishment of a properties-based joint standard for data transmission or schema and taxonomy formats, as well as the proposed properties. The Agencies also invite comment on whether, as an alternative, it would be preferable to establish specific data transmission and schema and taxonomy formats as joint standards. The Agencies also invite comment on use of the terms "data transmission format" and "schema and taxonomy format."

We generally support properties-based formats for data standards, and appreciate the distinction between the format for the data structure (schema/taxonomy) versus the format for the data packages (data transmission format). We recommend the agencies establish XBRL as the priority option for the schema/taxonomy format as the joint standard, and consequently the XBRL-compatible data transmission formats as the priority option for data transmission formats.

We recognize that properties allow for the most flexibility for adoption of newer formats in the future. While multiple formats may satisfy the proposed data standards properties, each format has different strengths and weaknesses. We believe that using established formats that are sufficiently expressive is very important in reducing the need to extend lesser standards and additional custom work to facilitate compliance and data quality.

As one of the main objectives of the FTDA is to promote interoperability of financial regulatory data, clear definition of data interoperability and the criteria to promote interoperability are of foundational importance. According to ScienceDirect, data interoperability refers to:

> *"...the ability to accurately interpret data that is exchanged between different systems or organizations. It involves ensuring that the data has* ***clear and unambiguous meaning****, is correctly mapped, and is formatted in the required form. Achieving data interoperability can be* ***challenging*** *due to semantic heterogeneity among models and* ***the need for data mediation to handle differences in naming, scales, and representations****."*

This is consistent with the FDTA law which requires that the data standards *"...(ii) enable high quality data through schemas, with accompanying metadata documented in machine-readable taxonomy or ontology models, which clearly define the semantic meaning of the data, as defined by the underlying regulatory*

information collection requirements." (§5811[2])

While multiple schema/taxonomy standards are available, having the same attributes represented across different formats with resolution to the data-point is paramount for maximizing data interoperability and its value realization. Of the schema/taxonomy formats discussed in the proposed rule, XBRL represents the richest model that is already designed for business reporting. Especially for reporting with extensible data points, XBRL has the longest history of wide adoption by regulators globally, and the most industry assets for enabling taxonomies, data quality, and processing.

Taxonomies across different authorities can be created in consistent ways to represent types of data elements, the expectations based on the different data types, additional metadata related to the disclosure, how the elements should be visualized in a presentation (human readable view), whether and how the elements may participate in a calculation relationship, etc. While XBRL is extensible and thus taxonomies do not necessarily all work exactly the same ways in every respect, the same types of taxonomies can generally be supported in similar fashion with low customization. With established adaptable platforms in the filing ecosystems, more taxonomy standards are supportable and scalable. To further facilitate and ensure scalability, we would also recommend boiling down the structured data reporting regimes to some main types, and align the different information collections to each based on criteria that makes a type of structured data regime more suitable.

XML, on the other hand, is less suitable for increasingly complex reporting requirements. In our experience, three main factors make XML a less efficient data standard.

1. XML has historically required more effort than XBRL to fully implement a data standard for business reporting. XML provides a flexible structure for encoding any type of data with no predefined tags, requiring users to define their own schema. The high flexibility can lead to unnecessary complexity.
2. The validation rules and structures also need to be specified in the schema each time in order to validate the data against the schema. If one were to reproduce in XML the structures already provided in XBRL, one would also need to write software to perform validation–which is already readily available for XBRL in open source and commercial solutions.
3. Finally, when a schema is used to support different types of entities with different requirements, the schema is typically too general and does not contain enough information on how each respective requirement should be supported and validated.

Due to the flexible nature of XML, there is much more room for XML schemas to be designed in different ways with different structures and elements representing the same data points and validated in different ways across data collections. This leads to higher difficulties in managing the experience across different reporting obligations for the regulated entities, and incremental efforts required to make the data truly interoperable. The

[2] https://www.congress.gov/bill/117th-congress/house-bill/7776/text

lack of specificity in the schema for varying requirements also results in additional custom development by providers to fully implement the data standards, potentially leading to scaling issues and cost burdens.

Moreover, while XBRL is developed based on XML, the acceleration of global XBRL adoption from 32 implementations worldwide in 2010 to 216 in 2023 (based on the XBRL project directory[3] information provided by the XBRL International) is evidence that XBRL is substantially more effective for regulators. We encourage the agencies to consider the case study published by the XBRL International on "XBRL or Customized XML"[4] for further information comparing the two standards.

In summary, we encourage the agencies to leverage a rich schema/taxonomy format like XBRL to standardize digital data collections as much as possible. We also encourage the agencies to take into consideration the longer lead time necessary for the filing ecosystems to implement certain formats (such as XML) as the agencies scale for wider adoption of data standards. And in the event that other formats like XML are necessary, we urge that the schemas should contain adequate specificity to enable proper functioning and validation, bringing consistent quality from all solutions providers. Where the schemas may not be the best place to provide such specificity, additional rulesets similar to the XBRL Data Quality Committee (DQC) rules utilized by the SEC would be very valuable.

Now we can turn to our comments regarding the data transmission formats.

Data packages can come in different flavors from the taxonomy/schema and can already take more than one output format. For example, XBRL (i.e., XBRL-XML) versus Inline XBRL, which are both accepted by the SEC and could be generated from the same XBRL taxonomy. A data transmission format is simply the syntax for the data files. Data can be structured using one taxonomy model but output into different data transmission formats for various reasons. For example, smaller file sizes may be important for certain types of reporting for faster data transfer.

XBRL does not only provide a rich format for structuring data, it also has an unique advantage in supporting the data output in different data transmission formats thanks to the XBRL International's Open Information Model ("OIM") that supports XBRL-XML, XBRL-JSON, and XBRL-CSV, allowing data packages to be rendered in JSON and CSV formats while maintaining full compatibility and resolution with the XBRL data model.

We agree that the data syntax should utilize living standards to avoid obsolescence, as we have seen data formats evolve over time and can be expected to continue to in the future. It would be best, however, that the agencies do not take on the burden of developing and maintaining these standards directly but to adopt them from the markets. Data standards should be evaluated in conjunction with the open source assets available in

[3] https://www.xbrl.org/the-standard/why/xbrl-project-directory/
[4] https://www.xbrl.org/guidance-files/XBRLorCustomizedXML.pdf

the ecosystem that complement the agencies' and the filing communities' needs. Consider, for example, the public company reporting in XBRL representing the most complicated structured data reporting at the SEC, is well-supported by the existing XBRL standards for both taxonomy and data transmission, and this did not burden the agency or the filing community with creating a foundational effort from scratch. We highly encourage the agencies to take this into consideration when assessing other data formats. PDF/A, for example, may only be fit for purpose in a highly constrained use case, and could yet create a burden on agencies to develop capabilities for data extraction.

Lastly, as further illustration of the superior function of XBRL versus XML for data transmission we can compare the visual rendering of their respective machine-readable documents. It is feasible to create an end-user friendly (i.e., non-technical) Inline XBRL viewer that works with any and all Inline XBRL documents without needing taxonomy-specific customization. In fact, there is an open source Inline XBRL viewer available for ready use by agencies and solutions providers, significantly reducing the burden of supporting the Inline XBRL standard (see the GitHub repository at https://github.com/Arelle/ixbrl-viewer). A viewer of XML data requires additional coding to create a non-technical end-user viewer for that data, and since XML is significantly less structured than XBRL, a viewer would need additional coding to support each and every distinct XML schema.

6. The Agencies invite comment on: (option 1) whether to establish a joint standard for taxonomies based on certain properties, and if so, the properties that should be set forth in the joint standard; or (option 2) whether to establish specific taxonomies, and if so, the taxonomies that should be set forth in the joint standard (such as those listed above or other specific taxonomies).

High data quality stems from data interoperability via taxonomy, a means of expressing standards. Joint standards for taxonomies are highly welcomed because they provide the same way of expressing standards across agencies, reducing silos or inconsistent overlaps. We recommend agencies continue the established practice of building small domain/topic-specific standards and promote reuse across information collections, which as previously discussed increases reporting consistency and data interoperability. This is preferable over a small set of large monolithic standards. Smaller, domain/topic-specific standards allow more nimble taxonomy creation, domain alignment across different teams, and refinement over time. For example, the Financial Accounting Standards Board (FASB) is responsible for the US GAAP taxonomy that is utilized by operating companies reporting financial information to the SEC, while other taxonomies for SEC specific disclosure requirements are maintained by the SEC, such as the Document and Entity Information (DEI) Taxonomy, Executive Compensation Disclosure (ECD) Taxonomy, Filing Fees Disclosure (FFD) Taxonomy, etc. These taxonomies can be reused across all domestic companies reporting to the SEC, and the regulated

entities gain much leverage from learning the taxonomies.

Taxonomy standards should focus on common data points. When multiple taxonomies are designed to be used together, guidelines should be developed to promote harmonization between the taxonomies. We recommend considerations in a few areas as follows:

1. When a taxonomy or a combination of taxonomies is designed to support more than one type of reporting entity, provide entry points that would bring in all the necessary content for the particular use case, and provide guidance and data validation requirements for each use case.
2. Promote consistency in taxonomy presentation so when multiple taxonomies are used together the navigational experience would be consistent between different portions of the taxonomy.
3. Ensure the usage of an element in one taxonomy does not conflict with the usage in another taxonomy.
4. Minimize exception handling, such as allowing or requiring modifications to the taxonomy where the reporting is closed and custom taxonomy is not expected.
5. Avoid mixing extensions and non-extension requirements in different parts of the reports or mixing other restrictions.

We recommend including these principles to guide the agencies in their agency-specific rulemaking efforts.

7. The Agencies also invite comment on use of the term "taxonomy" and whether the Agencies should define the term by rule, and if so, how the term should be defined.

The FDTA established significant requirements for the creation of the machine-readable data standards:

> *"(ii) enable high quality data through schemas, with accompanying metadata documented in machine-readable taxonomy or ontology models, which clearly define the semantic meaning of the data, as defined by the underlying regulatory information collection requirements."*

The FDTA references two terms of art from data science: taxonomy, and ontology.

> ***"taxonomy** is a hierarchical classification system that categorizes things based on their similarities and differences. It typically involves arranging items into groups and subgroups based on a set of criteria. Taxonomies are commonly used in fields such as biology, where organisms are classified according to their physical characteristics, and in information science, where taxonomies are used to organize and retrieve information."*[5]

[5] https://datafoundation.org/news/financial-data-transparency-hub/28/28-Implementing-the-FDTA-From-Data-Sharing-to-Meaning-Sharing

> ***"ontology** is a more complex system than a taxonomy that not only categorizes things but also defines the relationships between things. It is a formal and explicit specification of a conceptualization [...], that describes the entities, relationships and concepts that exist within a particular domain. Ontologies are commonly used in artificial intelligence to represent knowledge in a machine-readable format, and in the semantic web to enable machines to understand the meaning of data."* [ibid]

It may be helpful to distinguish these from "schema," a term also seemingly used in the proposed rule as interchangeable with "taxonomy."

> ***"schema** means, at its fundamental level, a specific plan or scheme. The schema of a data storage system (i.e., database) is primarily focused on a plan to logically and physically organize, store and retrieve data – not necessarily referring to the data's meaning nor facilitating any consistency of understanding it. Read "physically" as a computer database. This doesn't mean that data semantics cannot be represented (at least in part) in a schema but that the common practice is to divorce the model of meaning (an ontology) from the storage and retrieval model (a schema)."* [ibid]

The key point is not the selection of the term but the selection of the meaning, and this already is established in the FDTA. It uses the general term "schema" to mean a structure, and then it specifies that this structure needs to capture semantics (among other representation requirements), and that the means to represent these details are through "machine-readable taxonomy or ontology models."

We recommend a definition of "taxonomy" that forgoes the traditional definition of it being a hierarchical set of classifications that uses levels that go from broad categories to more specific ones. We do this because a dominant approach to creating semantic data standards includes the uses of non-hierarchical structures for organizing classifications (i.e., ontologies). If the joint rule wants to use the word "taxonomy" as a catchall then the definition should be broad as well:

> A taxonomy is a structured system for classifying and organizing concepts within a specific domain, ranging from simple glossaries to complex ontologies that define semantic relationships between concepts.

Other General Considerations

As one of the goals in the FDTA is to further enable AI applications, besides defining the term "Collections of Information" it is also important to note that level of data granularity could play a significant role. Data standards that support the datapoint level provide more information for regulatory insights and AI applications. AI requires good quality data input, and structured data is much better input than unstructured data source

such as 'plain' PDF, as when deriving structure from unstructured data, machines require extensive training and such models are typically not "nonproprietary" or "available under open license". We understand that some use cases of very short reports where fields are manually filled in with text may be suitable for PDF with embedded/named tags, permitting automated extraction with custom codes. However, in our experience, reporting entities welcome automation of data entry, data quality validation, and audit trails. Thus, the manual process of PDF forms, even with automated extraction, is counter to reducing the costs of report creation, and also counter to where technology providers (like Workiva) are heading.

While implementing structured data could, although not necessarily, impose a significant burden on the regulated entities depending on the type of structured data reporting, we have observed an average of 65% time reduction after the initial setup for the most complex form of structured data reports. Many financial agencies' reports are form-based, and automation of fixed forms is an even greater opportunity to reduce the initial and ongoing implementation burden. For example, the Federal Energy Regulatory Commission (FERC) XBRL implementation follows a fixed form approach and our XBRL conversion is highly automated with little need for manual intervention. Similar approach could be applied to a wide range of fixed forms, of small size or large, with a fillable form style of support. Regulated entities could then leverage technologies to manage their dataflow for structuring these forms. The result is an overall streamlined and efficient reporting experience that is otherwise hard to achieve without defined, consistent data standards.

In conclusion, we would like to once again express our appreciation for the joint data standards, and suggest the agencies to focus on data standards that provide the richest data model and the most open resources that complement these standards to minimize standards development, reuse of standards and implementation in taxonomy wherever applicable while maintaining harmonization, and strategies in the data standards implementation to lower costs of compliance while maximizing interoperability and information value of the machine readable data.

We would be pleased to further discuss our comments at your convenience should you have any questions.

Sincerely,

David Haila
EVP, Chief Technology Officer
Workiva Inc.