Data Standards & the Financial Data Transparency Act (FDTA)

May 2023

Sponsored by
Donnelley Financial Solutions

DFIN

Table of contents

Executive Summary

Senator Mike Crapo (R-ID) described the aims of the Financial Data Transparency Act (FDTA) [1], as follows, *"Making financial data used by federal regulators more accessible and understandable to the American public is an important step in improving government transparency and accountability."* [2] The FDTA, jointly introduced by Senators Crapo and Warner (D-VA), represents a real opportunity to meet these goals.

Implementing the right data standard, as called for in the FDTA, will enable economies of scale, reduce the cost of reporting, data collection and analysis, and generate good quality, actionable data for policy-setters, regulators, and the public, including investors, and researchers.

Standards like UPCs, QR codes and shipping containers, take an existing process or task, and improve its efficiency and effectiveness. Shipping containers, for example, have a standard, engineered structure and design, that optimizes the transport process - enabling automation, economies of scale, increased delivery speed and inventory fidelity (less theft and breakage). UPCs and QR codes track inventory and can take audiences to destinations without exposing details that might cause confusion.

Data standards have a similar purpose and impact. They take the guesswork out of communicating and transporting information which improves data reliability. They reduce human involvement in data processing and enable economies of scale through automation.

> *The FDTA, properly implemented, will allow regulators to see both the forest and the trees.*

Open data standards leverage a competitive marketplace of software tools, lowering the cost of reporting, data collection and analysis.

Where we are today

FDTA agencies maintain over 400 data collections from thousands of reporting entities in multiple formats including PDF, text, HTML, custom XML and XBRL. The current state of data processing and management among the agencies that fall under the FDTA negatively impacts reporting entities, regulators, and other data users. Users of data have limited access to machine-readable, interoperable data. Disclosure requirements are often fragmented and ambiguous. Data cannot be easily located, inventoried, or stored. Entity and securities identifiers are not consistently applied which makes it nearly impossible to effectively evaluate business and investment risk. Reporting entities face significant duplication in reporting, and confusion in contending with

numerous forms. Both reporting entities and data users are faced with lengthy technical documentation on how to report and use data, with no linkage between the data reported and the semantic data model. Today's approach has evolved over time, with each agency laser-focused on their own reporting needs. Not surprisingly, this has led to a highly siloed approach to data management which causes many of the problems outlined above. If regulators truly wish to reduce reporting burden, enable economies of scale, and encourage more timely, transparent reporting, they must coordinate efforts and work together.

What success looks like

As regulators work toward the plan to roll out the FDTA, it is critical to keep in mind what constitutes success: reliable, unambiguously machine-readable, interoperable data, a reduction in reporting burden and cost across all stakeholders, and adaptability to changes in reporting needs and technology over time.

Hundreds of effective data standards programs have been rolled out by regulators worldwide. U.S. based programs by the Federal Deposit Insurance Corporation (FDIC), and the Securities and Exchange Commission (SEC) were launched 18 and 15 years ago, respectively, and continue to expand because of their success. The Federal

Energy Regulatory Commission (FERC) initiated its first data standards program in 2021 and is working to expand on that program as well.

How to get there

The roadmap to effective data standards that will meet the letter and the spirit of the FDTA is already tested and proven in hundreds of programs worldwide. Agencies that fall under the FDTA have a clear path to follow:

Step 1: Build taxonomies (digital dictionaries) that unambiguously describe each data collection: reported facts, relationships between facts.
Step 2: Review data collections to eliminate duplicates and consolidate reporting needs.
Step 3: Consolidate reporting across all FDTA agencies to (again) eliminate duplicates and reduce reporting burden.
Step 4: Educate agencies, reporting entities, and intermediaries that support a robust, competitive reporting infrastructure.

The FDTA concretely stipulates data standards that "*render data fully searchable and machine-readable*," and that "*enable high quality data through schemas, with accompanying meta-data documented in machine-readable taxonomy or ontology models.*" Nevertheless, alternatives to data standards are likely to be considered. Options such as spreadsheets, custom XML schemas, and artificial intelligence may be contemplated.

While these approaches may be considered easier to implement, regulators must carefully consider not only the requirements of the legislation itself, but more importantly, the short- and long-term impact of each approach considered. These alternatives will not meet the requirements, nor will they meet the goals of the FDTA. Open data standards will.

> **Alternatives like spreadsheets, AI and custom XML will not meet the goals or requirements of the FDTA. Data standards will.**

This paper explores the current state of data management among agencies and provides a roadmap to meet the achievable goals laid out in the FDTA. The ability to link reporting requirements across agencies through universal data standards will give regulators, for the first time, a holistic view of regulated entities. The FDTA, properly implemented, will allow regulators to see *both the forest and the trees.*

Current State

FDTA agencies maintain 449 data collection programs based on our analysis of the Interagency Data Inventory[3], a spreadsheet prepared by the Data Committee of the Financial Stability Oversight Council (FSOC) which provides an inventory of data collections by FSOC member organizations. The breakdown by agency is shown below. We further classified

the collections in FSOC's inventory into financial and non-financial data sets, notifications, applications, surveys, and recordkeeping. The kind of information deemed to be a "collection" ranged widely, from requiring a bank to notify customers about various policies orally or by email, to a highly prescribed form, to the submission of financial statements. Appendix A provides a breakdown by agency and explanation of the categorization approach.

A variety of data formats are used to collect data including PDF, custom eXtensible Markup Language (XML), eXtensible Business Reporting Language (XBRL), text, Comma Separated Values (CSV), Word, and Hypertext Markup Language (HTML).

PDF is the most common format used for reporting. Preparation of a PDF form may require the reporting entity to manually enter data into the form (keyed in, or printed and written), save it, and submit it to the regulator as a PDF. The regulator receiving the data may need to manually extract the data from the PDF and key it into their financial system.

Alternatively, the PDF form may be structured with a schema that is able to automatically read information that has been keyed in as digital data and extract it into the regulator's financial system. Neither PDF-based solution is ideal: the first option requires manual data entry and extraction; the second requires building a

Board of Governors of the Federal Reserve	Consumer Financial Protection Bureau (CFPB)	Federal Deposit Insurance Corporation (FDIC)	Securities and Exchange Commission (SEC)	Municipal Securities (SEC)	Federal Housing Finance Administration (FHFA)	National Credit Union Administration (NCUA)	Office of the Comptroller of the Currency (OCC)
141	11	25	142	48	23	5	54

custom data extraction designed to fit the schema of the specific report. Custom schemas (discussed in greater detail later on in this paper) require custom tools, and data reported is not interoperable with other data reported using different custom schemas.

Financial data standards

There are many financial standards in existence today that are used for a variety of reporting needs as shown in a sample list on the table below. As FDTA agencies begin exploring the optimal approach, they should consider these standards. More detail is provided in Appendix B.

Messaging	FpML FIX Protocol ISO 20022 OFX
Financial Reporting	XBRL
Financial Securities Identification	CUSIP Number ISIN SEDOL FIGI Ticker Symbols
Entity Identification	LEI CIK DUNS OCC ID NFA ID Number FDIC Certificate Number RSSD BIC RTN/ABA UEI
Financial Instrument Classification (product)	CFI Code FIBO UPI
Other	ISDA SDR Reporting Standards EMIR Reporting Standards

The cross-agency review conducted identified inconsistencies in data collection approach from agency to agency, and in some cases, even within a single agency.

The lack of coordination between agencies has led to a siloed approach, which proper implementation of FDTA requirements has the potential to rectify. Below are observations and examples of the challenges for both reporting entities and users of data with the current process.

Problems for regulators & data users

Limited availability of machine-readable data

There are approximately 45 (some still in proposal stage) out of 449 data collections required or proposed to be submitted in XBRL or custom XML format, both of which can produce machine-readable data. XBRL is a data standard for the electronic exchange of financial and business data. It provides a common format for reporting financial information, such as financial statements and regulatory filings, in a machine-readable form. Among FDTA agencies, the SEC, Federal Reserve and FDIC collect data in XBRL.

While custom XML formats generate machine-readable data, they are unlikely to incorporate a broader framework where standardized and agreed upon concepts are used across data collections. Custom XML formats are typically created based on a specific reporting need.

The Federal government technology landscape is littered with single-use custom-developed applications for reporting.

Most other reports collected by FDTA agencies are submitted in PDF, which is a machine-readable document, but *on its own,* does not generate machine-readable data. PDF and other formats can be created with a schema which would support the creation of machine-readable data when a completed form is received by the regulator. This kind of custom schema however, is unlikely to incorporate a data standards framework because it is created based on a specific form. An LEI defined in one PDF form and an LEI appearing in another XML submission cannot be automatically identified as representing the same kind of fact because each form follows its own custom schema.

> *PDF files are machine-readable documents, but do not generate machine-readable data.*

Lack of interoperability

Data reported in non-machine-readable formats like PDF, text, Word, or Excel, cannot be commingled, shared, or inventoried together. XML data collections generate machine-readable data but cannot be automatically commingled with other machine-readable data

Proper implementation of the FDTA means:

- *Reliable, machine-readable data that can be automatically consumed and analyzed.*

- *Increased interoperability across agencies, with data that can be inventoried, easily queried and understood, shared, and transported.*

- *Reduced reporting duplication and burden.*

- *Adaptability to reporting changes over time inexpensively and easily.*

- *Facility to adjust to new technologies as the market evolves.*

because XML-formatted data is typically based on a schema custom-designed for a particular reporting situation.

For example, the SEC requires financial statement reporting from Regulation A, Regulation Crowdfunding, and publicly listed companies. Each type of company adheres to different reporting schemas as shown in the image of schemas below. Reg Crowdfunding and Reg A companies follow two different custom

```
Regulation Crowdfunding - Form C
<totalAssetMostRecentFiscalYear>382349.00</totalAssetMostRecentFiscalYear>

Regulation A - Form 1-A
<totalAssets>100.00</totalAssets>

Public Company Form 10-K
13de-40ac-9626-81be51fe5876" unitRef="usd">7266851000</us-gaap:Assets>
```

XML schemas. Public companies prepare their data in XBRL which is a different schema. The image shows the concept Assets formatted three different ways depending on the type of company and schema required. Form C uses the term "totalAssetMostRecentFiscalYear", Form 1-A uses "totalAssets", and corporate filers reporting on 10-K or 10-Q, use "Assets". Not only is there a different concept to represent the same kind of fact, but the term used for Form C is defined differently to include the time period reported.

Filing agents and other providers that serve SEC filers often work with many types of SEC reporting companies. Under the current scenario, a provider that serves all three types of companies (Reg A, Reg CF, and public companies) must develop and maintain three separate products which is more costly than supporting a single application for all. If a single schema were used for all reporting applications, the cost could be shared across companies of many types, enabling economies of scale that benefit reporting entities in the form of lowered costs from providers.

Fragmented disclosure requirements

Banks are required to report to the SEC and the FDIC, but can opt out of SEC disclosure requirements by filing Form D, notice of an exempt offering of securities with the SEC. While this reduces the reporting burden, it can have unforeseen consequences.

Signature Bank, which closed its doors on March 12, 2023, is one of the banks that opted for just such an exemption. Signature is also part of the S&P 500 which makes it of particular importance to equity investors.

Despite that, the bank's data was not easily accessible to the equity markets. Signature Bank is the second S&P 500 bank that opted out of SEC reporting. Republic Bank, which was also seized shortly after Signature Bank and sold to JPMorgan Chase, also opted for the exemption[4]. An inconsistent regulatory approach makes it especially challenging for equity analysts to get a real picture of the market as a whole, and banks in particular.

Authors of the study, *Fragmented Securities Regulation, Information-Processing Costs, and Insider Trading*[5] addressed inconsistencies between SEC and FDIC bank disclosures, noting, "*Our findings suggest regulatory fragmentation adversely affects the market efficiency and level playing field by increasing information-processing costs, a novel mechanism through which regulatory fragmentation creates costs to the financial system.*"

Ambiguities in data

Forms, like the FHFA Community Support Statement shown below, sometimes do not provide clear instructions defining the data to be reported. For example, Part II, Section A, highlighted in the red box, requires the reporting of the number and dollar amounts of loans, as well as ratios. It is not clear either in the form or

FEDERAL HOUSING FINANCE AGENCY
COMMUNITY SUPPORT STATEMENT

FHFA Form # 060
(01/10/2012)

(see instructions page 2)

Name of Institution:_____

Address: _____

City:_____ State:_____ Zip Code: _____

FHFA ID Number: _____

Contact Person: (Mr./Ms.)_____Title: _____

Phone Number: _____   Email: _____ **or** Fax: _____

Part I. Community Reinvestment Act (CRA) Standard.

Most recent federal CRA Rating:_____   CRA Evaluation Date: _____

Part II. First-time Homebuyer Standard: *All members must complete either Section A or B. Members with "Outstanding" federal CRA ratings need not complete this part. Members should use data or activities for the previous or current calendar year.*

A. Complete the following four questions: *If your institution did not track loans or made no loans to first-time homebuyers, you must complete Section B.*

1. Number of mortgage loans made to first-time homebuyers   # _____
2. Dollar amount of loans made to first-time homebuyers $ _____
3. Loans made to first-time homebuyers as a percentage of all mortgage loans   _____ %
4. Dollars loaned to first-time homebuyers as a percentage of all mortgage dollars loaned   _____ %

---

the accompanying instructions, however, over what time period these figures should be calculated. This hinders the ability to aggregate data, review data over time, or conduct meaningful comparisons between institutions.

Multiple entity identifiers, proprietary securities identifiers

A single organization may have multiple, incompatible legal entity identifiers such as a Central Index Key (CIK) for SEC reporting, a Unique Entity Identifier (UEI) for those obtaining federal funding, a Bank Identifier Code (BIC), a Data Universal Numbering System (DUNS) number, the Replication Server System Database ID (RSSDID) used by the Federal Reserve, or an Employer Identification Number (EIN). The plethora of entity identifiers reduces the ability to reliably match, track and monitor

business and investment risk. This, coupled with the fact that many organizations are composed of hundreds or thousands of legal entities, makes it unnecessarily difficult to associate reported data with a specific entity and the ultimate impact of that data on a holding company.

The most common securities identifier in use today in the United States is the CUSIP (Committee on Uniform Securities Identification Procedures). CUSIP is owned by the American Bankers Association and operated by FactSet Research Systems Inc. Because the CUSIP is commercially owned, its usage can be subject to licensing fees.

The proprietary nature of the CUSIP identifier restricts the ability to link data within the

financial system. The uncertainty surrounding copyright and a litigious position taken by CUSIP in the past results in suboptimal alternatives being used to the CUSIP such as stock tickers. Tickers identify the exchange on which a security is traded, but they are not consistently defined across exchanges and tickers do not exist for many securities.

The proprietary nature of the CUSIP also causes confusion in the marketplace as to its appropriate use. The European Commission required that S&P divest CUSIP services due to market dominance concerns[6], and recent cases in the US create uncertainty around the appropriate use of the identifier[7].

The Financial Instruments Global Identifier (FIGI) is an alternative securities identifier that is freely redistributable and non-proprietary. FIGI is an Object Management Group (OMG) standard that was developed by Bloomberg and made publicly available through OMG. FIGI is still administered by Bloomberg. The advantage of this identifier is that it is non-proprietary, global, and covers most security types. It also allows identification of securities by the exchange on which they are traded. None of the other existing security identifiers that are in widespread use, are global, cover all major security types, are non-proprietary and include the markets on which the security is traded. The FIGI effectively covers the work done by CUSIP, ISIN, SEDOL, VALERON tickers and numerous other standards used in smaller markets.

Regulators globally, with the exception of Brazil, have been reluctant to adopt the FIGI standard. The SEC only recently allowed the use of the FIGI on the Form 13-F (an SEC form that lists securities holdings of the registrant) but they continue to require the use of the CUSIP on Form 13-F, even if the issuer uses the FIGI.

Problems for reporting entities

Duplicate reporting requirements

Publicly listed banks report the same data to both the SEC and the FDIC, although the data reported follows different accounting standards for each regulator. The figure on the next page shows FFIEC Form 041 for MainStreet Bancshares compared to their 10-Q report for the same period. Data is often represented differently and, in some cases, has slight variations because of rounding differences.

Schedule RC - Balance Sheet(Form Type - 041)

All schedules are to be reported in thousands of dollars. Unless otherwise indicated, report the amount outstanding as of the last business day of the quarter.

Dollar amounts in thousands

1. Cash and balances due from depository institutions (from Schedule RC-A):		
a. Noninterest-bearing balances and currency and coin[1]	RCON0081	17,998
b. Interest-bearing balances[2]	RCON0071	86,735
2. Securities:		
a. Held-to-maturity securities (from Schedule RC-B, column A)[3]	RCONJJ34	15,170
b. Available-for-sale debt securities (from Schedule RC-B, column D)	RCON1773	161,819
c. Equity securities with readily determinable fair values not held for trading[4]	RCONJA22	0
3. Federal funds sold and securities purchased under agreements to resell:		
a. Federal funds sold	RCONB987	0
b. Securities purchased under agreements to resell[5]	RCONB989	0
4. Loans and lease financing receivables (from Schedule RC-C):		
a. Loans and leases held for sale	RCON5369	0
b. Loans and leases held for investment	RCONB528	1,461,066
c. LESS: Allowance for loan and lease losses	RCON3123	12,994
d. Loans and leases held for investment, net of allowance (item 4.b minus 4.c)[7]	RCONB529	1,448,072
5. Trading assets (from Schedule RC-D)	RCON3545	0
6. Premises and fixed assets (including capitalized leases)	RCON2145	21,295
7. Other real estate owned (from Schedule RC-M)	RCON2150	0
8. Investments in unconsolidated subsidiaries and associated companies	RCON2130	0
9. Direct and indirect investments in real estate ventures	RCON3656	0
10. Intangible assets (from Schedule RC-M)	RCON2143	0
11. Other assets (from Schedule RC-F)[8]	RCON2160	108,723
12. Total assets (sum of items 1 through 11)	RCON2170	1,859,813

On FFIEC 041 Cash and Cash Equivalents is reported as the sum of Noninterest-bearing and Interest-bearing. On the 10-Q, it is reported in total.

Total assets on the two forms match with a slight rounding error.

MAINSTREET BANCSHARES, INC. AND SUBSIDIARY

Consolidated Statements of Financial Condition as of September 30, 2022 and December 31, 2021 (Dollars in thousands, except share and per share data)

	At September 30, 2022 (unaudited)	At December 31, 2021 (*)
Assets		
Cash and due from banks	$ 50,636	$ 61,827
Federal funds sold	54,098	31,372
Cash and cash equivalents	104,734	93,199
Investment securities available-for-sale, at fair value	162,319	99,913
Investment securities held-to-maturity, at amortized cost	17,670	20,349
Restricted securities, at cost	16,436	15,609
Loans, net of allowance for loan losses of $12,994 and $11,697, respectively	1,448,071	1,341,760
Premises and equipment, net	14,523	14,863
Other real estate owned, net	—	775
Accrued interest and other receivables	8,273	7,701
Bank owned life insurance	36,996	36,241
Computer software, net of amortization	7,258	2,493
Other assets	43,835	14,499
Total Assets	$ 1,860,115	$ 1,647,402

Identical forms used separately by different agencies
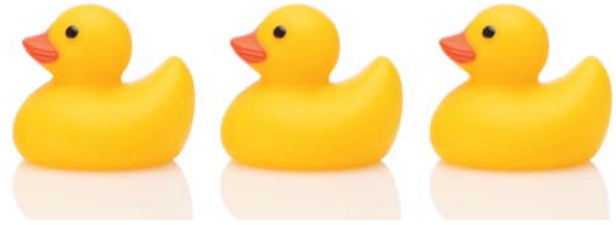
SEC Form 4, Statement of Beneficial Ownership, is identical to FDIC Form 4, Statement of Beneficial Ownership of Securities, as shown below. Similarly, the SEC and FDIC maintain two versions of Form 3, Initial Statement of Beneficial Ownership, and two versions of Form 5, Annual Statement of Beneficial Ownership of Securities. Banks report these forms to either the SEC or the FDIC, depending on the number of shareholders in the bank. Consolidating these six forms into three, with submission to a single portal where data can be pulled by both agencies, would streamline the data collection process, and improve the usability of the data by making data from all reporting entities available in a single location and structure.

Problems for all stakeholders

Preparing and using data requires extensive documentation review

Many reports require lengthy explanatory documentation before data preparation and reporting can begin. For example, the Federal Reserve form, Financial Statements of U.S. Nonbank Subsidiaries of U.S. Holding Companies, FR Y-11, is a 9-page PDF form that contains financial and text data as shown on the image below. The instruction document for this forms is 58 pages long, containing detailed definitions on who should file, when and where, along with detailed definitions for each reporting concept. The right side of the image below shows the General Instructions for the income statement portion of the report.

Forms and instruction documents can become outdated. Multiple versions may be in use and in

distribution which could result in inconsistencies in reported data and market confusion.

Similarly, many publicly available data sets are accompanied by technical documentation that must be reviewed to understand and use the data files. FHFA for example, publishes loan-level data from Fannie Mae and Freddie Mac in text files with a corresponding dictionary in PDF format which assists users in understanding the data files. Below is a sample row in one of the text files. Each row in the file represents one single-family property.

2   5 1 2 2 2 1 8 4 5 9 1 5 1 4

The characteristics of the property outlined in the text file can only be understood by reading lengthy documentation. The partial report image below shows the meaning of the first four data fields on the row. A data user reads the custom designed data model in the technical documentation in order to consume this text file into its financial system. For example, the mort-

gage on the property defined in the row to the left was purchased by Freddie Mac (field 1 is reported as value 2), has record number 5, is in a metropolitan area (1), and has a percent minority between 10% and 30% based on the 2010 Census Tract. Because there is no automated linkage between the data model and the data reported, it is cumbersome, inefficient, and costly to report, extract, and analyze the data. Documentation may not be kept up-to-date or outdated versions may still be in distribution. Furthermore, using this loan-level data requires building a custom program to extract and interpret the reported data.

Another example illustrating the challenges to data users is found in Appendix B featuring credit union data.

Enterprise Public Use Database
Single-Family Properties
National File A

RELEASE OF 2021 DATA

The "National File A" contains mortgage-level data on owner-occupied 1-unit properties.
Note: Fields are separated by one blank space.

Field #	Field Width	Field Name	Values	Description / Comments
1	1	Enterprise Flag	1 = Fannie Mae 2 = Freddie Mac	Flag identifying whether the mortgage was purchased by Fannie Mae or by Freddie Mac. Fannie Mae and Freddie Mac are collectively referred to as the Enterprises in this document.
2	7	Record Number		Sequential numerical identifier for the property not related to the record number in the Census Tract File or the other National Files.
3	1	Metropolitan Statistical Area (MSA) Code	1 = metropolitan area 0 = non-metropolitan area	Location of the property based on the MSA definitions in effect on January 1, 2021.
4	1	2010 Census Tract - Percent Minority	1 = >=0, <10% 2 = >=10, <30% 3 = >=30, <=100% 9 = Missing	The percentage of the census tract's population that is classified as belonging to a minority group, based on the 2010 decennial census.

Future State

A well thought-out data standards program is a long-term solution that will continuously adapt to market needs and changes. Proper implementation of the FDTA means:

- Reliable, machine-readable data that can be automatically consumed and analyzed.
- Increased interoperability across agencies, with data that can be inventoried, easily queried, and understood, shared, and transported.
- Reduced reporting duplication and burden.
- Adaptability to reporting changes over time, inexpensively and easily.
- Facility to adjust to new technologies as the market evolves.

To reach the desired goals, FDTA agencies must adopt the appropriate data standards, entity identifiers and securities identifiers. Data standards and identifiers must be globally oriented. Many of the entities required to report to FDTA regulators, from public companies to banks, operate in a global marketplace. Requiring US-centric standards, like CUSIP, would limit the benefit across these entities as they are likely to need to adhere to global standards if they are listed in non-US markets, or conduct business in other countries.

A "standard" that is only used by one or two market participants, is not a standard. Standards and identifiers must have broad market support, be widely used, and be supported by many software vendors.

The challenge facing the FDTA agencies today is to conduct the initial work to fully understand their own reporting needs by building taxonomies that fully represent all data collected.

Commonly used standards like UPCs and shipping containers provide their greatest benefit as support becomes more widespread. Like the loading equipment that supports the expansion and use of the shipping container, software vendors that support a data standard are equally important. More vendors mean greater competition, and downward pressure on costs across the supply chain.

Standards and identifiers must be flexible and adaptable to change, and in turn, must ease the process for stakeholders to adapt to change as well. For example, regulators must be able to change reporting requirements (request new information to be reported, or revise existing requirements), and it must also be easy for reporting entities to efficiently transition to these requirements.

Data standards must be capable of handling unique reporting requirements but structured enough to unambiguously capture whatever is reported in a consistent manner. Data standards shared among the FDTA agencies must be capable of transforming financial, narrative, and other data types into digital information that is unmistakably machine-readable.

Taxonomies developed to support the FDTA must be modularized. The FDTA covers many reporting entities, many types of reported data, and multiple regulators. A modular taxonomy design will allow pieces of a taxonomy to be built independently and inserted into the taxonomy structure where needed. Individual agencies can maintain their own unique reporting requirements but share those that cross across more than one agency.

A single concept may be re-used hundreds of times in multiple reports or financial statements. An individual report is sometimes re-used by more than one agency. A modular approach will allow the flexibility to handle requirements that may cross over from one agency to another, while at the same time reducing duplication in reporting.

And finally, standards and identifiers chosen must be open-source and non-proprietary as required in the legislation. This important characteristic will ensure the lowest possible cost for regulators, reporting entities, and data users.

Proven method to build data standards

The assessment of data collections addressed earlier in this paper is only the start of the process to build and adopt data standards required by the FDTA. To bring these disparate collections together using uniform data standards will require evaluating each data collection in much greater detail. This assessment can be tedious and painstaking, involving consultation among regulators, reporting entities, analysts, and researchers, as well as software providers that are involved in preparation or data extraction. There is no shortcut to building data standards that yield the desired results. Once the upfront work is completed, however, the way forward is easier, more cost-effective, and productive for all members of the supply chain.

The FDTA calls for data standards that "*(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.*"

A taxonomy is a means of classifying information or objects into categories, for example breeds of dogs, or types of plants.

A Taxonomy or Ontology representing data is a hierarchically-arranged digital dictionary of terms representing everything that must, or can, be disclosed for a particular reporting situation. It also describes the relationships that exist between data. An XBRL Taxonomy is the single representation of the data model which is an abstract illustration of the organized data elements and their relationships based on real-world objects.

The information contained in the Taxonomy should seek to define the data model as completely as possible. This approach develops the taxonomy as the "single source of truth" or the "Single Data Model." It means that the Taxonomy contains everything needed by the data collector, the reporting entity, the data intermediary, software applications used to generate reports or analyze data, and data consumers. The Single Data Model eliminates the need for separate instructional materials, or documents containing definitions, properties or the agreed upon presentation of reported facts.

Everything is available and accessible in the Taxonomy, which is referenced by applications used by stakeholders to report, collect, extract, and analyze data.

Advantages of this approach include:
- Ensures that everyone has the most current reporting requirements. Reduces the chance of using outdated forms or instructions.
- Increases efficiency and reduces the reporting burden, because reporting applications and the entities that use them will digitally reference the current requirements through the Taxonomy.
- Improves data consistency because the single data model ensures all participants have access to the most current documentation and definitions for concepts.
- Enhances efficiency and ease of change in reporting requirements when needed. The regulator (or manager of the Taxonomy) can add, revise, or delete concepts *once*, in the Taxonomy (single dictionary of terms) and the change is automatically communicated to all.
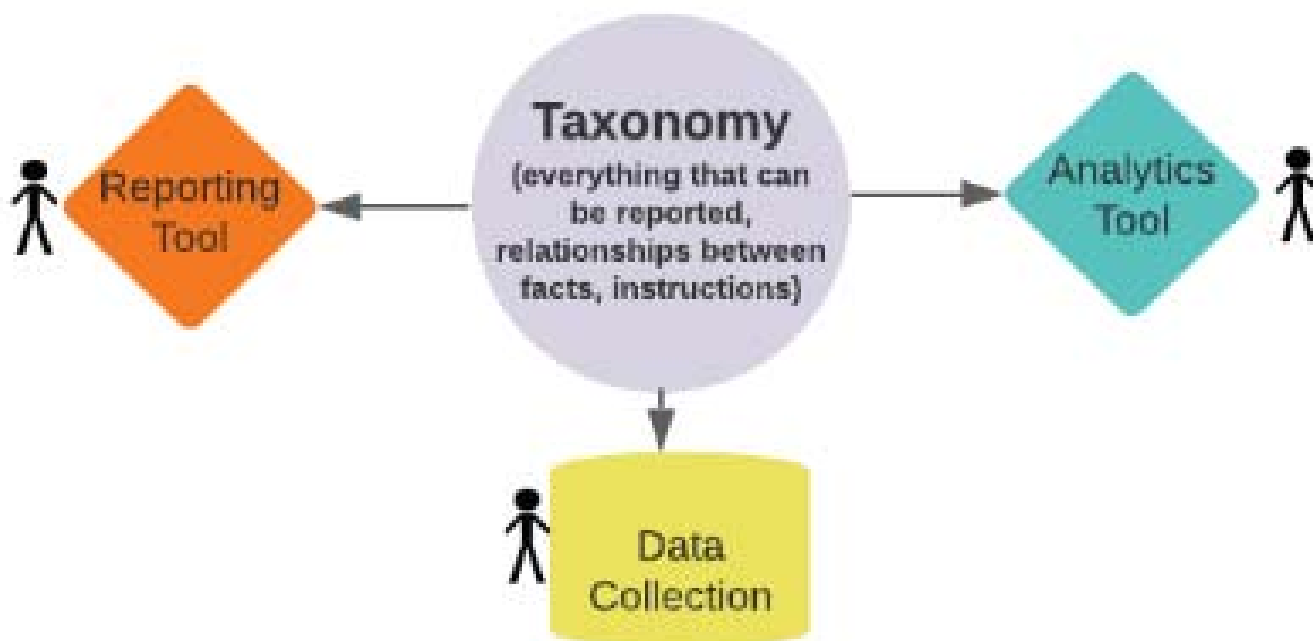
The visual above shows the Taxonomy, as the single data model, which is referenced by all stakeholders whenever they need to report, collect, or extract information.

When data is prepared using a Taxonomy (the data model), it is created using a technical format that transports data structured in conformance with the data model. Commonly used formats used to convey data reported in a standardized data model include XML, JSON, CSV, and XHTML. As shown in the visual below, the Single Data Model (Taxonomy) provides the semantic meaning of the data and the format (XML, XHTML, CSV, JSON) provides the transport to make the data portable.
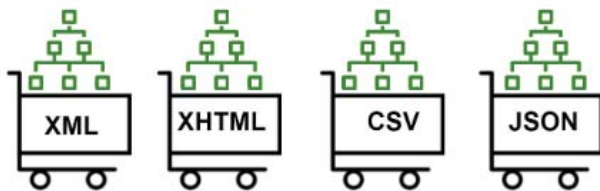
Step 1. Build taxonomies.

The first step in building data standards is for each agency to identify the data collections to be standardized, and create taxonomies for each collection. The agency starts by concretely defining each fact on a report or statement, regardless of whether the data is made publicly available or is strictly used for internal agency purposes. Every reported fact is associated with a specific data field or concept, which will have additional attributes. The visual below illustrates how a fact is reported on the Federal Reserve's FR Y-11 Schedule IS-B. It has associated properties which can include a computer-readable name, a human-readable label, data type, period type, balance type, definition, and authoritative references. All properties shown in the blue bubbles will be codified in the taxonomy.

Schedule IS-B—Changes in Allowance for Credit Losses[1]

Dollar Amounts in Thousands	(Column A) Loan and Leases		(Column B) Held-to-Maturity Debt Securities[2]		(Column C) Available-for-Sale Debt Securities[2]	
	BHCS	Amount	BHCS	Amount	BHCS	Amount
1. Balance most recently reported at end of previous calendar year (i.e., after adjustments from amended Income Statements)	3124		JH88		JH94		1.
2. Recoveries	4605		JH89		JH95		2.
3. LESS: Charge-offs	C079		JH92		JH98		3.
4. Provision for credit losses[3]	4230		JH96		JH96		4.
5. Adjustments	4815		JH91		JH97		
6. Balance at end of current period (sum of items 1, 2, 4, and 5 minus item 3) (must equal Schedule BS, item 3.b)	BHCT / 3123		JH93		JH99		6.

Definition = loans and leases held for investment.

Name = RecoveriesOnLoansAndLeases

Label = Recoveries on Loans and Leases

Data type = monetary

Period type = duration

Balance = credit

References = FFIEC Schedule IS-B

Reported concepts can also have relationships to other concepts. Three standard relationships used in XBRL are:

Presentation. A representation of the accepted ordering of concepts on a report or financial statement, for example, assets are followed by liabilities, then by equities. In the example above, the concept Allowance for Credit Losses of Loans and Leases appears in column A at the top of the table, to represent beginning balance, and at the bottom of the table, to represent ending balance. For concepts like this, the period of the associated fact determines where the concept to be associated with each fact appears in the presentation.

Table (definition). A financial statement balance sheet has line items representing concepts that are allowed to be reported on a balance sheet, like assets and liabilities. The same table does not allow income statement concepts like revenues or expenses, to be reported as line items on the balance sheet.

Calculation. Concepts may be mathematically related, for example, Current Assets is a child to Assets, and a sibling to Noncurrent Assets.

Other relationships can also defined between concepts. All information: relationships between concepts, as well as properties, labels, and references for each concept, must be captured in the Taxonomy (single data model).

Considering the FHFA Community Support Statement again, we look at the characteristics of the fact that would appear in the red box. The characteristics of the data fields are classified the same way as the data fields on the Federal Reserve form even though the data is quite different.

Certain properties, like balance type, do not apply to every data field. Properties that every concept must have are a computer-readable name, human-readable label, definition, period type, and data type.

The third example shown on the next page is the NCUA Call Report Form 5300. Again, a concept on this report can be captured by identifying the same characteristics as with the other reports. This example shows a concept that has mathematical relationships with other concepts. Coin and Currency is a child to Cash on Hand, and to Total Cash on Deposit (item d) which represents the sum of Cash on Hand and Cash on Deposit. These relationships must also be captured in the taxonomy.

As you can see from the repetitive nature of the labels, properties and references associated with each concept in these three examples, this process can be followed quite easily for each data collection. Identifying relationships between concepts within each data set can also be understood by exploring how the facts are presented and used.

Step 1. Build taxonomies.

Step 2. Individual agency review and consolidation.

Step 3. Cross-agency review and consolidation.

Step 4. Educate and implement.
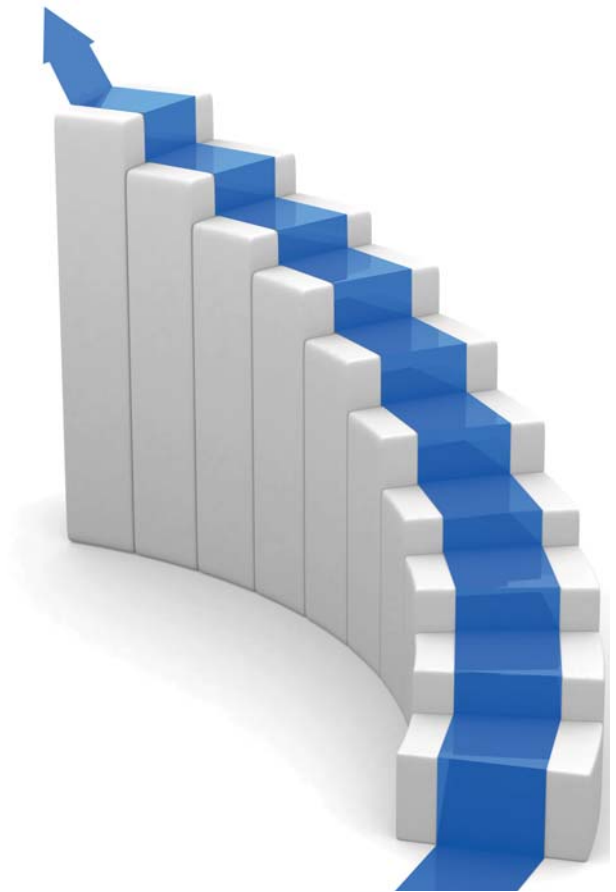
Credit Union Name: _____

Federal Charter/Certificate Number: _____

STATEMENT OF FINANCIAL CONDITION AS OF: _____

This page must be completed by all credit unions.

ASSETS

	Account
Have you adopted ASC Topic 326: Financial Instruments - Credit Losses (CECL)? Select yes or no.	AS0010
NOTE - Review the Call Report Instructions carefully if you have adopted ASC Topic 326: Financial Instruments - Credit Losses (CECL).	

CASH AND DEPOSITS: If your credit union reports an amount in Account AS0007 complete Schedule B, Section 3, Investments - Maturity Distribution.

		Amount	Account
1. Cash on Hand	a.	Coin and Currency	AS0004
	b.	Cash Items in Process of Collection	AS0005
	c.	Total Cash on Hand	730A
2. Cash on Deposit (Amounts Deposited in Financial Institutions)	a.	Cash on Deposit in Corporate Credit Unions	730B1
	b.	Cash on Deposit in a Federal Reserve Bank	AS0003
	c.	Cash on Deposit in Other Financial Institutions	730B2
	d.	Total Cash on Deposit (Amounts Deposited in Financial Institutions)	730B
3. Time deposits in commercial banks, S&Ls, savings banks, natural person credit unions, or corporate credit unions			AS0007
4. All other deposits			AS0008
5. TOTAL CASH AND OTHER DEPOSITS (Sum of Accounts 730A, 730B, AS0007, and AS0008)			AS0009

Name = CoinAndCurrency

Label = Coin and Currency

Data type = monetary

Period type = instant

Balance = debit

References = NCUA Call Report Form 5300

Definition = Amount of coin and currency on hand.

The result for each agency will be a clearly defined set of taxonomies representing each agency data collection (report). The set of FDIC (red) and SEC (green) taxonomies shown below are identically structured but each one represents a different data collection. There will likely be duplicate concepts in each taxonomy which can be consolidated into a set of "base elements" for each agency. For example, FDIC Taxonomy Report 1 and FDIC Taxonomy Report 2 may both require the reporting of Assets. That concept can be put into a base FDIC Taxonomy and shared across reporting entities. That takes us to the next step.

Step 2. Individual agency review & consolidation

After each agency has data consistently structured and defined, they will evaluate their set of taxonomies set with an eye towards eliminating duplicate concepts, and refining definitions, labels, and references.

In some cases, the agency may decide that one concept meets the needs of multiple reporting requirements. The agency works with stakeholders to obtain agreement on labels and definitions. In other cases, it may be determined that two data fields that appear to be the same, have subtle differences, and therefore the definitions and labels may require further refinement to better articulate the differences. Those concepts that are common to two or more reports, should be pooled together into a set of "common concepts."

This process will result in each agency establishing a base taxonomy with common concepts and a set of smaller taxonomies that represent additional data fields that are unique to a single data collection. As shown below, a reporting entity who uses SEC Taxonomy Report 1 will use the Report 1 entry point to locate the concepts needed. Report 1 will access the SEC Common Taxonomy to bring in more concepts that are shared with others to complete the reporting requirements for Report 1. An "entry point" is designed to give the reporting entity only those concepts he or she needs for a specific reporting situation, to make it easier to prepare the report.

Step 3. Cross-agency review & consolidation

This step mirrors step 2 but involves all FDTA agencies. The agencies together will conduct a cross-agency evaluation of all data collections. Again, duplicates will be identified, and data fields that appear similar but may be used for different purposes (and therefore are not the same) will be further defined to articulate subtle but important differences.

A base FDTA taxonomy composed of common concepts will be created that every agency can draw from for their separate agency data collections. As illustrated on the right, an FHFA reporting entity accesses the FHFA Taxonomy for Report 2, locates some concepts in FHFA Report 2, some in FHFA Common Taxonomy, and some in the Base FDTA Taxonomy. The latter contains all those concepts that are common, and therefore shared, across all agencies.

This modular approach is useful because some entities report to more than one FDTA agency and may even report the same information to two separate agencies.

It is also useful for those situations where taxonomies may need to be maintained by outside organizations. For example, today, the SEC relies on taxonomies created by both the Financial Accounting Standards Board (FASB) and the International Accounting Standards Board (IASB). The former maintains the US GAAP Taxonomy; the latter the IFRS Taxonomy. This is a logical approach because these organizations are the accounting standard setters and best equipped to ensure that the taxonomies can model the most current accounting standards. The SEC manages additional taxonomies that are used in conjunction with the FASB and IFRS taxonomies.

When the FDTA rolls out, the SEC is required to establish data standards for municipal bond issuers as well. Those standards would logically be maintained by the Governmental Accounting Standards Board (GASB) which is part of the Financial Accounting Foundation (FAF), just like the FASB. The figure below illustrates how the SEC would be able to maintain these multiple reporting requirements for different entities.

The issuer on the left is an IFRS filer. He or she enters the SEC IFRS Taxonomy which references both the IFRS Taxonomy in brown and the SEC Common Taxonomy. The SEC Common Taxonomy references the Base FDTA Taxonomy for those data fields that are common to all FDTA reporting, for example, legal entity identifiers, organization name, and certain commonly used financial concepts such as Assets. The filer can locate all the concepts needed to accurately represent the IFRS financial statements and document information.

Similarly, the SEC filer who follows US GAAP on the top of the image enters the SEC US-GAAP Taxonomy which references the FASB Taxonomy and SEC Common Taxonomy. The latter, in turn, pulls Base FDTA Taxonomy data concepts.

And finally, the municipal bond issuer on the left side of the image enters the municipal taxonomy which references the GASB maintained taxonomy, as well as the SEC Common Taxonomy and Base FDTA Taxonomy data concepts.
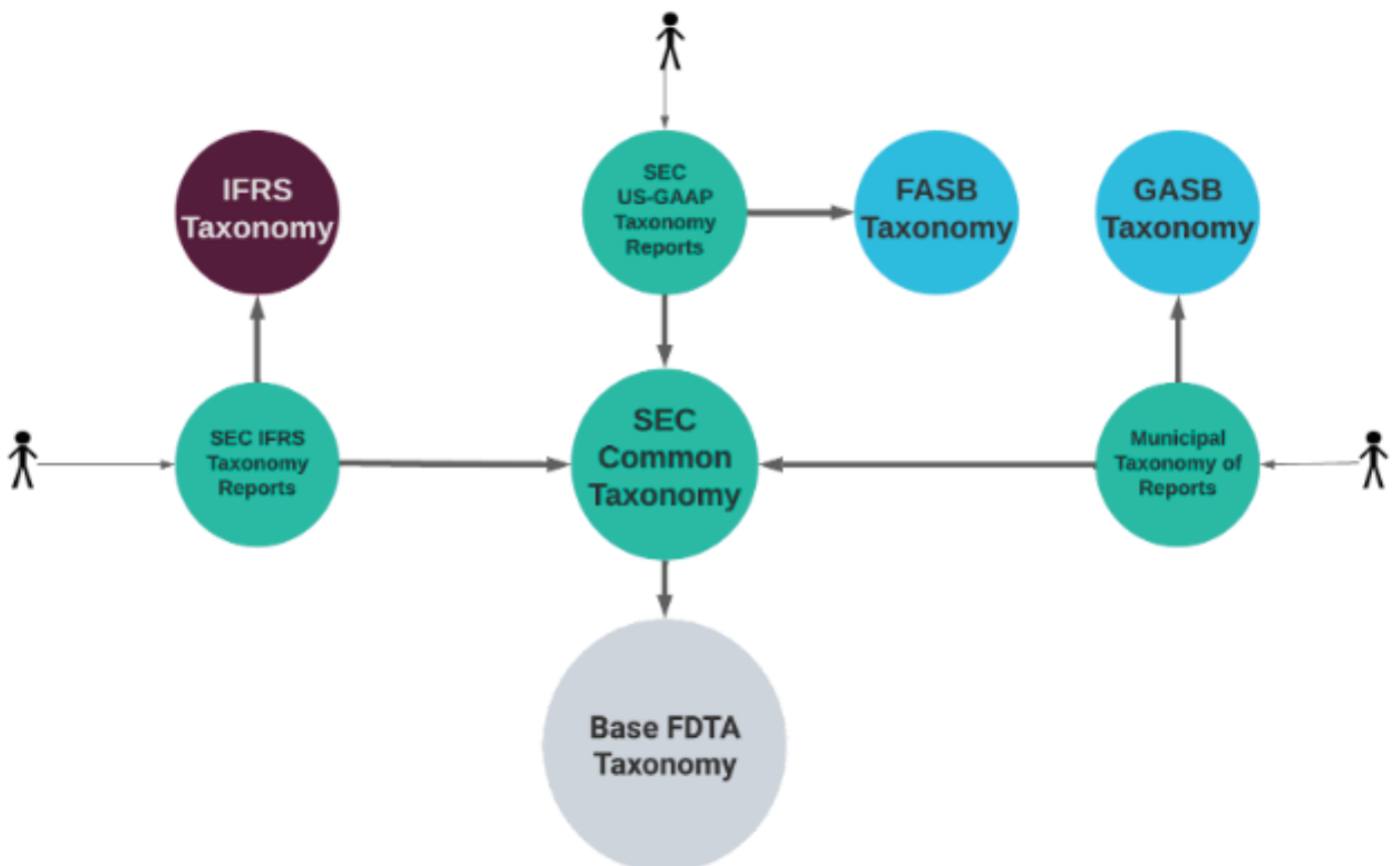
Step 4. Educate and implement.

For a successful FDTA implementation, data reported must comply with relevant accounting standards, and agency regulatory requirements. This information is contained in the agency's common and collection-specific taxonomies. The data is made available either solely to regulators or to the public at large and must also be in compliance with the taxonomies, including the FDTA Taxonomy of cross-agency concepts.

Each agency will control their own agency-specific taxonomies and can opt to mandate the currency in which data is reported, frequency of reporting, and what data must be submitted, just as they do today.

Agencies can also choose the formatting technology used to transport the data. Formatting technologies, for example, could be XML, CSV, JSON, or XHTML. For example, one agency may opt to have data reported to them in CSV format. A separate agency may wish to have data reported in XML files. The consistency across data sets is in how the data is structured and defined - the semantic data model. The flexibility of data transport ensures that the data standards program can adapt to new technology formats that may become available over time.

The program will require significant education and training to the stakeholder community, to assist regulators, reporting entities, data users, and software providers that support the reporting and use of data.

Options to implement

The FDTA legislation calls for the use of open, non-proprietary data standards that can render data searchable and machine-readable, with financial reporting requirements documented in machine-readable taxonomies or ontologies. Standards adopted must incorporate standards developed and maintained by voluntary consensus standards bodies and be consistent with applicable accounting and reporting principles. Below are options that may be under consideration as regulators explore FDTA implementation, although not all will satisfy the objectives or the requirements of the legislation.

Spreadsheets

Developing a spreadsheet of data fields with associated properties for each concept is the first step in developing data standards. A spreadsheet alone, however, is not a data standard. It cannot represent relationships between facts. It cannot render data machine-readable.

Data reported in a spreadsheet cannot be structured in a standard format that off the shelf software can reliably extract from unless

the spreadsheet is highly structured and consistent. Even when prepared in a very prescribed format, software reading the spreadsheet can easily misinterpret reported data, and it can be nearly impossible to identify what is wrong. Furthermore, data required to be reported can often exceed the available space allowed on a single sheet.

AI/machine-learning

Artificial Intelligence (AI) has many potential applications, but it lacks the accuracy and reliability of data standards. AI can assist in the standards development process but on its own is not a replacement.

In theory, regulators could publish general rules defining the data required to be reported and allow reporting entities to provide the data in unstructured format such as PDF, as many do today. The unstructured data could then be consumed and categorized by an AI engine to generate structured, machine-readable data.

This approach, however, requires very large data sets, as well as significant computing power and storage capacity. Organizations like regulators that wish to consume this data, will need to invest in high-performance hardware and infrastructure, and cultivate AI expertise to develop, train and test the AI model. There are no economies of scale with this approach because every organization that wishes to consume the data will bear the same costs to

establish their AI platform. Different AI platforms may generate different results, producing inconsistent data from user to user.

Alternatively, the regulator could opt to manage a single AI platform and generate machine-readable data which is then provided to all, but this still incurs significant costs, and requires very specific expertise. If the regulator is responsible for generating the structured data set, it also bears the liability for the accuracy of the data rather than the reporting entity. Even more importantly, AI has not progressed to the stage where it can produce dependable, consistent financial data.

AI's ability to learn patterns, however, has the potential to improve the way data standards are created by automating many of the tasks involved such as identifying concepts that should be incorporated into a taxonomy, and normalizing as-reported data. Accounting standards bodies, regulators, and investors can use AI to quickly identify trends and risks in reported data, particularly if that data is structured to begin with, that can then be evaluated and incorporated into the next release of the data model (taxonomy).

Custom XML

eXtensible Markup Language (XML) is a flexible data format that can embed or "tag" information in a reported fact that renders it understood by reporting entities, intermediaries, and end users. Data prepared using a custom XML schema reporting financial statement data, for example, could be developed that captures the definition, label, time period, and other characteristics of the fact in a single value reported for a fact like Assets. The schema could be represented by something like this:

```
<Assets>
    12100
</Assets>
```

Because of the flexibility of XML, an alternative XML schema could be built that could represent the same fact along with the time period for the fact, like this:

```
<AssetsMar_31_2022>
    12100
</AssetsMar_31_2022>
```

A third XML schema could be created to represent the same data as this:

```
<Assets03_31_2022>
    12100
</Asset03_31_2022>
```

The data model explaining the meaning behind the fact 12,100, is built into each XML file that is reported. No separate taxonomy is needed. XML is commonly used, and many software applications can easily create and consume XML-formatted data. XML is effective at creating portable, machine-readable data.

XML generates machine-readable data but not in a standardized data format. Custom XML produced data is not interoperable with other data sets.

However, XML is not a data standard. Opting forXML would require creating individual custom schema for each data collection that accurately represents time period, tabular data, reporting entity, etc. Each schema is likely to be structured differently from others because XML gives agencies the flexibility to build whatever they want. This *will* generate structured data, but *not in a standardized data format.*

The absence of a defined structured format for data created means that separate tools used to report, collect, extract, and analyze data would need to be developed for every custom data collection.

Making changes to reporting requirements through a custom schema is costly across the supply chain, because it requires every tool and every system supporting the reporting, collecting and use of the data, to be updated to adapt to the change.

Data prepared in custom XML would be required to be reported in XML, limiting the use of other technologies like JSON, CSV, XHTML, or technology formats that may be introduced in the future.

XBRL

eXtensible Business Reporting Language (XBRL) has the same ability to render data machine-readable and portable as XML. Unlike XML, XBRL has a concretely defined structure that consistently communicates information about reported data like time period, units, data type, dimensional characteristics, definitions, and relationships between reported facts.

XBRL is an open, non-proprietary global data standard used for financial and business information reporting worldwide (213 implementations worldwide[8]). It has a consistent structure for the reporting of financial and business information that can be used across all types of data collected across the FDTA agencies. It is already used by entities reporting to the SEC, the FFIEC (Federal Reserve and FDIC), and by the Federal Energy Regulatory Commission (FERC), although the latter is outside of the scope of the FDTA.

XBRL uses taxonomies to establish a consistent structure for reported facts which is used to render data fully searchable and machine-readable. It has the unique capacity to capture financial, narrative, and many other data types. It harmonizes with existing accounting and report standards like US GAAP, IFRS, and FDIC call reports, and can support other financial and non-financial documents. Because it is based on a taxonomy (single data model), regulators can make changes easily and inexpensively with XBRL, and reporting entities can adapt to

updates with minimal disruption. Public companies transition to updated taxonomies each year to report to the SEC. Banks transition to updated taxonomies, often each quarter to report to the FDIC.

XBRL today can be transported in multiple technology formats to give regulators flexibility in how they wish the data to be reported. XBRL International, the global voluntary consensus standards body that supports and maintains the technical specification, has an established program called the Open Information Model (OIM) [9], which focuses on adapting the specification to technology changes. Today, the specification is designed such that XBRL documents can be prepared in:

- XML, a file format to store, transmit and reconstruct data
- XHTML (eXtensible HyperText Markup Language), a file format that combines XML and HTML to render information both human- and machine-readable
- JSON (JavaScript Object Notation), a common file format with a simplified syntax used to store and convey data
- CSV (Comma Separated Values), a plain text file format, most useful for large volumes of consistently prepared data

XBRL International's OIM initiative ensures that the specification will continue to be expanded going forward to adapt as new technologies emerge. As shown on the visual below, the OIM has successfully adapted to new technologies as reporting needs evolve over time. The SEC, for example, required XBRL in XML format when their structured data program for corporate issuers began in 2009; in 2018, they transitioned to XHTML (called Inline XBRL) because SEC 10-K and 10-Q filings need to be both human- and machine-readable.

Inline XBRL was also mandated by the European Securities and Markets Authority (ESMA) for entities issuing securities in the European Union. This format works well as a PDF replacement, because it provides both machine-readability and the ability to produce visually attractive reports.

The XBRL-CSV data standard is most appropriate for large volumes of consistently prepared data. The SEC Form N-MFP for example, would be a good candidate. This form requires the reporting of repetitive rows of weekly and daily data for series-level and class-level information about a fund. Today, Form N-MFP is required to be reported in custom XML. Transitioning to XBRL-CSV would streamline and reduce the size of the reported file, because the data model is referenced in the taxonomy, rather

XBRL International Open Information Model (OIM)

An initiative to provide new and simpler ways to work with XBRL data by defining the meaning of data and allowing it to be transported by varying formats.
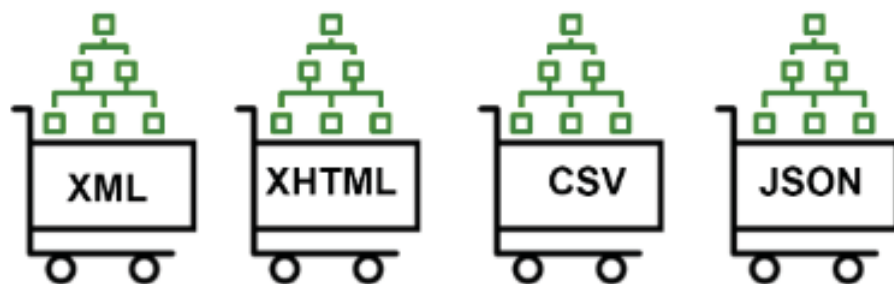
XML	XHTML	CSV	JSON
Introduced 2000	Introduced 2011	Introduced 2022	Introduced 2022

than in the file itself. The XML file below is a Form N-MFP. This illustrates the repetitive nature of the XML structure required. XBRL-CSV would produce machine-readable data just like custom XML, but it would render the data in data standard format and produce a smaller, more manageable file and the reported data would be in structured format that would be interoperable with other data sets.

```
▼<totalValueDailyLiquidAssets>
    <ns3:fridayDay1>21441505528.16</ns3:fridayDay1>
    <ns3:fridayDay2>19111194106.75</ns3:fridayDay2>
    <ns3:fridayDay3>18768568022.81</ns3:fridayDay3>
    <ns3:fridayDay4>18150096551.02</ns3:fridayDay4>
    <ns3:fridayDay5>19784728037.74</ns3:fridayDay5>
  </totalValueDailyLiquidAssets>
▼<totalValueWeeklyLiquidAssets>
    <ns3:fridayWeek1>23275532350.02</ns3:fridayWeek1>
    <ns3:fridayWeek2>22795454905.61</ns3:fridayWeek2>
    <ns3:fridayWeek3>22268125272.80</ns3:fridayWeek3>
    <ns3:fridayWeek4>21649788551.01</ns3:fridayWeek4>
    <ns3:fridayWeek5>23618190281.08</ns3:fridayWeek5>
  </totalValueWeeklyLiquidAssets>
```

XBRL is modular, allowing multiple taxonomies to be created independently and consolidated.

XBRL is developed and maintained by XBRL International, a global voluntary consensus standards body with robust technical working groups that continuously develop the standard to meet market needs. Because of the widespread use of the XBRL standard, there are thousands of software applications and tools, both open-source and commercial, that support the smooth efficient flow of data in XBRL format.

The XBRL standard is not as well-known or widely used as XML, CSV, or other formats. However, the OIM ensures that developers with experience in XML, JSON, and other formats, can easily work with XBRL.
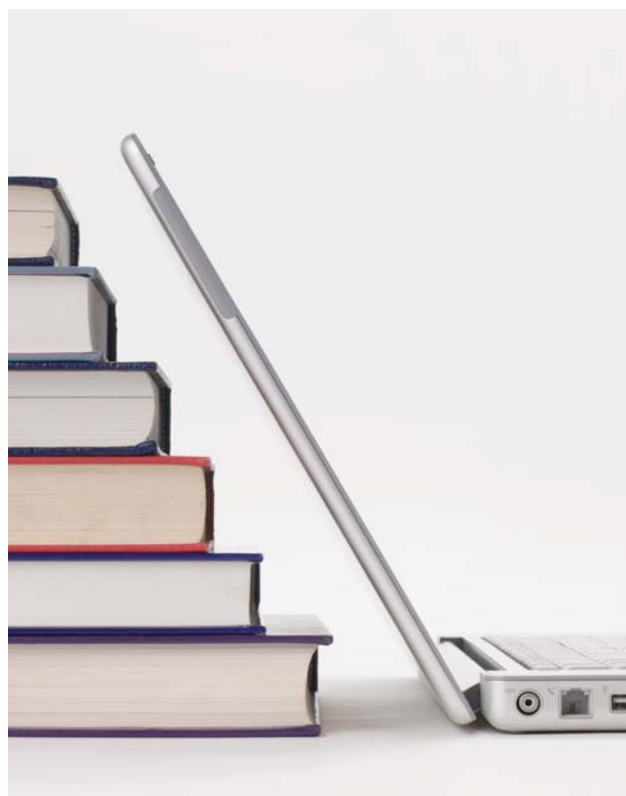
Conclusion

The current siloed, paper-based approach to data collection by financial regulators is inefficient and highly burdensome to regulators and to reporting entities.

The FDTA poses a rare opportunity to modernize reporting practices through automation. Properly implemented data standards programs are undeniably effective at reducing data processing cost, enabling economies of scale, and producing information that is timely and actionable. Programs like these in the US have met or exceeded the goals of the regulators and are being expanded to assist with more reporting requirements.

Reporting burden declines with the use of standards because duplication is eliminated. The cost of preparing data in standardized format has declined since data standards programs in the US began in 2005 with the FDIC, and in 2009 with the SEC. As standards, including data standards, become more widespread, costs will continue to decline even more.

The challenge facing the FDTA agencies today is to conduct the initial work to fully understand their own reporting needs by building taxonomies that represent all data collected.

"... we have regulators all around the world, including more than 10 in the United States alone. Regulations include stress testing, reporting, compliance, legal obligations and trading surveillance, among others.

While the business is the first line of defense on all these issues, we also have 3,700 people in compliance, 7,100 in risk and 1,400 lawyers actively working every day to meet the letter and the spirit of these rules along with the final line of defense — audit."

— Jamie Dimon, Chair & CEO
JP Morgan 2022 Annual Report

Appendix

A. Interagency Data Inventory Review

The Interagency Data Inventory Review spreadsheet lists 541 data collections associated with the FSOC agencies that fall under the FDTA. To determine which collections would be most appropriate for standardization, we eliminated certain categories of collections, including those deemed to be "one-time," some that appeared to be duplicates or that had been discontinued or replaced by another form. There were also a few that referenced various regulations but could not be defined as a true data collection. 449 collections remained after culling the list.

Agencies may determine that some of these collections are not strong candidates, for example, collections used solely for recordkeeping, or letter-based notifications to consumers about required policies and practices. Surveys with variable topics and/or respondents may also not be appropriate.

This review is based on investigating forms, instructional materials, and data sources online where available, and on consultation with various subject matter experts in certain areas of reporting. Conclusions drawn are based on a high-level review only. Further investigation by each agency may come to different conclusions.

	Federal Reserve	CFPB	FDIC	SEC	Municipal (SEC)	FHFA	NCUA	OCC	Totals
Datasets									
Financial	40	7	5	69	48	17	2	14	202
Non-financial	7	2	2	33		2	3	5	54
Applications									
Financial	2			4					6
Non-financial	13		5	13				10	41
Notifications									
Financial	5			4				1	10
Non-financial	32		10	16				5	63
Other collection types									
Surveys	22	2	3	1		3		8	39
Recordkeeping	20			2		1		11	34
Total	141	11	25	142	48	23	5	54	449

Each collection was reviewed and categorized into:

- Data collections, either financial or non-financial. This classification generally applied to periodic disclosures of financial or narrative information such as financial statement data, which may or may not be reported in a form. Examples of non-financial data sets include lists of bank directors, or registration to become a crowdfunding portal.
- Notifications, financial or non-financial. Notifications are often triggered by an event. Non-financial notifications may be the submission of a letter to a regulatory authority, or alerting customers to credit policies. Examples of financial notifications may include providing rate data to customers or notification to a regulator from funds seeking to deregister.
- Applications, financial or non-financial. Applications are often prepared as online PDF forms. Examples of financial applications include Application for Federal Reserve Bank Stock.
- Surveys and recordkeeping. Many surveys varied by topic or respondent and therefore may be a less likely candidate for data standardization. Most recordkeeping collections are also not prime candidates for standardization.

Observations about specific agency collections:

- FDIC.
  - Financial data collections that generate machine-readable data include XML and XBRL (Summary of Deposits).
  - Notifications are made to customers and to regulators.
  - Three regulator notifications are forms-based. All others are prepared letters or documents.

- Federal Reserve.
  - Of 22 surveys, many vary by topic or by target audience. Of those that are more targeted and/or consistent from period to period, some may be useful to collect in standard data format, for example the Survey of Small Business and Farm Lending. Some surveys are required to be submitted by financial institutions, for example the General Use Prepaid Card Survey and the Interchange Transaction Fees Survey - the volume of data and consistency make these appropriate for data standardization.
  - Most non-financial notifications are submitted as letters or documents, the balance are forms-based. Some are notifications to consumers, for example concerning credit policies or disclosures required to be made prior to finalizing a consumer lease.
  - Eleven out of 13 non-financial applications are forms.
  - 30 out of 40 financial data collections are on forms. FFIEC Forms 031, 041 and 051 are reported in XBRL.

- SEC.
  - Of the 69 financial data collections, 41 have XBRL or XML requirements, or are in proposal stage.

- Seven of the 33 non-financial data collections are in structured XML format or are in proposal stage. One is in proposal stage to be prepared in XBRL.
  - Of the 17 applications, 1 is prepared in XML.

- Municipal (SEC).
  - The Municipal Securities Rulemaking Board (MSRB) collects 48 report types into the Electronic Municipal Markets Access (EMMA) system which ranges from Primary Offering to Continuing Disclosures, all in PDF format.

- NCUA.
  - Two of the financial data collections are Call Report financial statement data; the third is data collected for credit union service organizations. The two non-financial data collections are corporate profiles.

- CFPB.
  - Data collection count for CFPB is 111 in total but the majority are one-time collections. Data collected for compliance with the Home Mortgage Disclosure Act (HMDA) on loan-level data is the largest dataset. It is collected in pipe delimited format and made available to the public in CSV and pipe delimited format.

- FHFA.
  - Membership of the Banks collection consists of a large number of spreadsheets used to track names, location, type, and multiple identifiers for 6,000 banks. Each year is represented in a separate spreadsheet which is difficult to connect to track bank changes over time. The Housing Mission Goals database contains a large volume of mortgage and housing sales by state and is posted in Excel spreadsheets.

- OCC.
  - Some financial and non-financial data collections are submitted in forms, some without a form. Non-financial notifications are typically made to consumers and/or regulators. Surveys are mostly non-financial.

B. Financial data standards

This is a partial list of data and identify standards. Other standards may be available beyond this list.

Messaging standards

FpML

Financial Products Markup Language (FpML) is a messaging standard used for the electronic communication of over-the-counter (OTC) derivatives data. It provides a common messaging format for the lifecycle of a trade, including trade capture, confirmation, and valuation.

FIX Protocol

The Financial Information eXchange (FIX) Protocol is a messaging standard used for the electronic communication of securities trading data. It is widely used in equities, fixed income, and foreign exchange markets. FIX Protocol allows market participants to send and receive real-time trading information, including trade orders and execution reports.

ISO 20022

ISO 20022 is a messaging standard used for the electronic exchange of financial data between different systems and institutions. It provides a common language and structure for financial messages, allowing for greater interoperability between systems and easier implementation of new payment and securities processing systems.

OFX

OFX is a messaging standard used to exchange financial data between financial institutions and personal financial management software.

Financial reporting standards

XBRL

eXtensible Business Reporting Language (XBRL) is a standard for the electronic exchange of financial and business data. It provides a common format for reporting financial information, such as financial statements and regulatory filings, in a machine-readable form.

Statistical reporting standards

SDMX

SDMX, which stands for Statistical Data and Metadata eXchange is an international initiative that aims at standardising and modernising ("industrialising") the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries.

Financial security identification standards

CUSIP Number

CUSIP stands for Committee on Uniform Securities Identification Procedures. A CUSIP number is a unique identifier assigned to financial instruments such as stocks, bonds, and other securities. It is used to identify securities for trading, settlement, and regulatory purposes.

ISIN

International Securities Identification Number (ISIN) is a unique identifier assigned to securities, including stocks, bonds, and other financial instruments. It is used for trading, clearing, and settlement purposes.

SEDOL

Stock Exchange Daily Official List (SEDOL) is a seven-character code used to identify securities listed on the London Stock Exchange and other UK exchanges. It is used for trading, clearing, and settlement purposes.

FIGI

Financial Instruments Global Identifier (FIGI) is a unique identifier assigned to financial instruments, including stocks, bonds, and other securities. It is used for trading, regulatory reporting, and risk management purposes.

Ticker Symbols

A ticker symbol is a unique series of letters assigned to publicly traded companies and their securities. It is used to identify securities for trading and tracking purposes.

Entity identification standards

LEI

Legal Entity Identifier (LEI) is a 20-character code used to identify legal entities, including corporations, banks, and investment funds. It is used for regulatory reporting and risk management purposes. The LEI maintains authorartative mapping tables between the LEI and ISINs (CUSIP) and LEI and BIC identifiers.

CIK

Central Index Key (CIK) is a unique identifier assigned to public companies and their filings with the US Securities and Exchange Commission (SEC). It is used for regulatory reporting and compliance purposes.

DUNS

The DUNS number is a nine-digit identifier that is assigned to a business or organization by Dun & Bradstreet (D&B). It stands for "Data Universal Numbering System". The DUNS number is based on a standardized numbering system that identifies the location, name, and industry classification of a business entity. The number is assigned to a single business entity and does not change, even if the business changes its name or location.

Entity identification standards, continued

OCC ID

OCC ID An identifier assigned by the Office of the Comptroller of the Currency (OCC) to all institutions that the OCC grants a charter. Beginning July 21, 2011, the OCC began using the OCC Charter field to facilitate storage of supervisory information for nationally chartered thrifts. These IDs consist of the unique identification number assigned to national banks supervised by the OCC to identify and track a bank. The OCC charters, regulates and supervises all national banks.

NFA ID Number

The National Futures Association (NFA) ID Number is a unique identifier assigned to firms and individuals who are registered with the NFA as commodity futures and options market participants. The number is used to identify market participants in regulatory filings and public disclosures.

FDIC Certificate Number

The Federal Deposit Insurance Corporation (FDIC) Certificate Number is a unique identifier assigned to banks and savings institutions that are insured by the FDIC. The number is used to identify banks in regulatory filings and public disclosures, and to track the deposit insurance coverage of individual accounts.

RSSD

Replication Server System Database (RSSD ID) is used by the federal reserve for identifying Banks.

BIC

The BIC (Bank Identifier Code) identifier, also known as the SWIFT BIC or SWIFT code, is a unique identification code that is assigned to financial institutions, including banks, in order to facilitate international wire transfers and other cross-border transactions. The BIC identifier consists of 8 or 11 alphanumeric characters and is structured in a way that enables the code to identify the name and location of the financial institution.

RTN/ABA

RTN/ABA Routing Transit Number (RTN) is a nine numerical digit number assigned to institutions by the American Bankers Association (ABA). The RTN number is commonly referred to as an ABA number and is found on the bottom of checks.

Financial instrument classification (product)

CFI Code

The Classification of Financial Instruments (CFI) code is a standardized code that identifies the specific type of financial instrument issued by an entity. The code is composed of six characters, with each character representing a different attribute of the instrument, including the asset class, the geographic region, and the type of instrument.ystemic risk.

Financial instrument classification (product), continued

FIBO

The Financial Industry Business Ontology (FIBO) is a standard vocabulary for representing financial concepts and relationships. It is a collaborative effort between industry groups and standards bodies to create a common language for the financial industry, facilitating the exchange of data and interoperability between systems.

UPI

The Unique Product Identifier (UPI) is a standard identifier for reporting over-the-counter (OTC) derivatives transactions to trade repositories. The UPI provides a unique code for each product, allowing regulators to track the overall market and assess systemic risk.

Other specific data collection standards

ISDA

The ISDA Standards refer to a set of standardized legal documents and definitions that are widely used in the derivatives market to establish the terms and conditions of transactions between parties. These standards include the ISDA Master Agreement, which sets out the basic terms and conditions for all transactions between parties, and various schedules, definitions, and confirmations that can be customized to reflect the specific terms of each transaction.

SDR Reporting Standards

SDR Reporting Standards refer to the rules and guidelines established by the Commodity Futures Trading Commission (CFTC) for swap data reporting to Swap Data Repositories (SDRs). The standards cover the data elements to be reported, the format of the data, and the timing of the reporting.

EMIR Reporting Standards

EMIR Reporting Standards refer to the guidelines established by the European Securities and Markets Authority (ESMA) for the reporting of derivatives transactions under the European Market Infrastructure Regulation (EMIR). The standards cover the data elements to be reported, the format of the data, and the timing of the reporting. ESMA utilizes a ISO 20222 message format, developed explicity for reporting between OTC counterparties and selected trade repositories.

C. Case Studies

Using credit union data

Credit unions report quarterly call report data to the NCUA just like banks that submit call report data to the FDIC. Below is an example of one page in the 32-page call report form for Hanscom Federal Credit Union, based in Bedford, Massachusetts.

The Form 5300 Call Report Quarterly Data files containing credit union financials are saved in a comma delimited text format that can be imported into a database or spreadsheet application such as Microsoft Access or Excel. The downloaded zip files contain over 20 text files that separately represent various sections of the call report or explanatory information.

We will consider a scenario where a data user wishes to extract data for all or some subset of credit unions for specific facts reported on Form 5300. To extract a fact such as "First mortgage loans sold on the secondary market" which is highlighted in red on the form above, the data user finds the "736" Account number in the AcctDesc text file which is contained in the zip file to identify the appropriate data table which in this case is the text file FS220B. The user then searches file FS220B to identify the field with the account number, 736. To match the 736 values with the name of the associated credit union, the user then needs to match the CU_NUMBER on the FS220B text file with the CU_NUMBER in a separate text file called FOICU.

Credit Union Name: HANSCOM   Federal Charter/Certificate Number: 9095

Back to Navigation Page

SCHEDULE A
LOANS, SUPPLEMENTAL INFORMATION AS OF: MARCH, 31 2023

SECTION 5 - INDIRECT LOANS

Complete this section if the credit union has any indirect loans outstanding. Report all indirect loans reported on page 6 regardless of acquisition method. Indirect loan participations must also be reported in Schedule A, Section 6.

		Number	Account	Amount	Account
1.	New and Used Vehicle Loans	17,783	IN0001	$245,200,616	IN0002
2.	First Lien and Junior Lien Residential Loans	0	IN0003	$0	IN0004
3.	Commercial Loans	0	IN0005	$0	IN0006
4.	All Other Loans	0	IN0007	$0	IN0008
5.	TOTAL OUTSTANDING INDIRECT LOANS (Sum of each column)	17,783	617A	$245,200,616	618A

SECTION 6 - LOANS PURCHASED AND SOLD UNDER 701.22 AND 701.23

LOANS PURCHASED		Year-to-date				Outstanding			
		Number	Account	Amount	Account	Number	Account	Amount	Account
1.	Loans Purchased from Other Financial Institutions	0	SL0014	$0	SL0015	0	SL0018	$0	SL0019
2.	Loans Purchased from Other Sources	0	SL0012	$0	SL0013	0	SL0020	$0	SL0021

LOANS SOLD		Year-to-date				Outstanding			
		Number	Account	Amount	Account	Number	Account	Amount	Account
3.	Loans Sold	0	SL0022	$0	SL0023				
4.	First mortgage loans sold on the secondary market	5	SL0024	$1,082,000	736				
5.	Loans Transferred with Limited Recourse Qualifying for Sales Accounting	0	SL0026	$0	819				
6.	Real Estate Loans Sold with Servicing Retained	5	SL0028	$1,082,000	SL0029	2,180	SL0030	$475,782,626	779A
7.	All Other Loans Sold with Servicing Retained	0	SL0032	$0	SL0033	0	SL0034	$0	SL0035

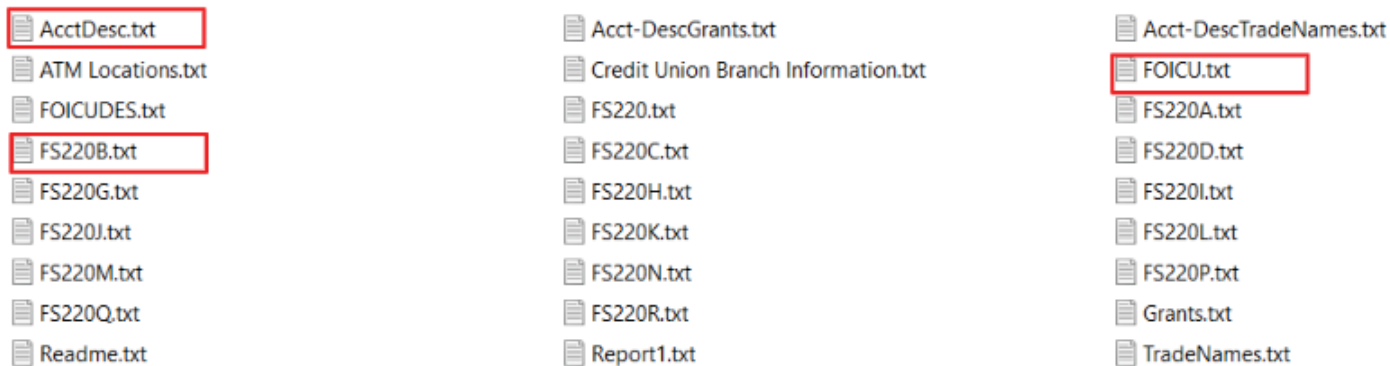| LOAN PARTICIPATIONS		Participations Purchased				Participations Sold			
		Outstanding Balance	Account	Amount Purchased Year-To-Date	Account	Retained Balance Outstanding	Account	Amount Sold Year-To-Date	Account
8.	Vehicle - Non-commercial	$0	SL0036	$0	SL0037	$0	SL0038	$0	SL0039
9.	Non-Federally Guaranteed Student Loans	$0	691L7	$0	SL0041	$0	691N7	$0	SL0043
10.	1- to 4-Family Residential Property	$0	691L2	$0	SL0045	$0	691N2	$0	SL0047
11.	Commercial Loans excluding Construction & Development	$9,264,059	691L8	$0	SL0049	$0	691N8	$0	SL0051
12.	Commercial Construction & Development	$0	691L9	$0	SL0053	$8,468,606	691N9	$0	SL0055
13.	All Other	$0	SL0056	$0	SL0057	$0	SL0058	$0	SL0059
14.	TOTAL (Sum of each column)	$9,264,059	691L	$0	690	$8,468,606	691N	$0	691

The visual below shows the text files in the zip file. Those highlighted must each be used to pull the data required by the data user in this scenario.

This is the process a data user will follow to create a data set of first mortgage loans sold on the secondary market by credit union, for a particular quarter. This data extraction can be managed programmatically, so that a data user can set up a process to pull the same data on a recurring basis. It is important to note, however, that a custom program must be created for each data extraction. Presumably, a data user that is interested in credit union data, may also be interested in data from other financial institutions which are likely to be reported in a different fashion.

Creating custom data extraction programs for every data collection is unnecessarily burdensome, time-intensive, and inefficient. The time spent creating the custom data extraction program is multiplied across all the data users that wish to retrieve some portion of this data.

Leveraging the same structured data standards for credit unions, banks, and other reporting entities would dramatically improve efficiencies.

Furthermore, banks report call report data in XBRL format and it is likely that they report many of the same concepts as credit unions. Adapting the FDIC Call Report Taxonomy to work with credit union call report data should be explored as an option.

AcctDesc.txt
ATM Locations.txt
FOICUDES.txt
FS220B.txt
FS220G.txt
FS220J.txt
FS220M.txt
FS220Q.txt
Readme.txt

Acct-DescGrants.txt
Credit Union Branch Information.txt
FS220.txt
FS220C.txt
FS220H.txt
FS220K.txt
FS220N.txt
FS220R.txt
Report1.txt

Acct-DescTradeNames.txt
FOICU.txt
FS220A.txt
FS220D.txt
FS220I.txt
FS220L.txt
FS220P.txt
Grants.txt
TradeNames.txt

Developing standards for government financial reporting

An estimated 90,000 government entities exist in the United States; approximately 30,000 of them have active municipal bond programs, and therefore are required to submit information to the Municipal Securities Rulemaking Board (MSRB) in the Electronic Municipal Markets Access (EMMA) system, a web platform that collects and provides information about municipal bonds, prices and market trends. Data reported to EMMA includes Official Statements, Advance Refunding Documents, Variable Rate Security Documents, continuing disclosure documents, financial and operating filings, and asset-backed securities filings. Most data, including the Annual Comprehensive Financial Report (ACFR), and the Official Statement are made available to consumers in PDF documents.

In 2018, the State of Florida signed legislation requiring the use of XBRL for local government reporting. At the same time, XBRL US established a working group of municipal securities analysts, public sector academic researchers, think tanks, software companies, and standards development experts, to explore how machine-readable standards could be developed to accommodate ACFR reporting. Observers to the working group include industry groups representing government entities such as the National Association of State Auditors, Treasurers and Comptrollers (NASACT) and the Government Finance Officers Association (GFOA), as well as the Governmental Accounting Standards Board (GASB), and U.S. Census.

The mission of the working group is to address the difficulties encountered by researchers

and analysts to aggregate data and compare financial performance of governmental entities because of the limitations of the PDF format. The working group aims to address this problem by designing schemas and XBRL implementation(s) for open data reporting of state and local government actual financial results.

Taxonomy development methodology

The approach taken to build effective, practical data standards was similar to what is proposed for FDTA agencies. While developing financial statement standards for government entities is smaller in scope, there are thousands of government entities, of all shapes and sizes.

First, we ensured that we had subject matter experts involved with deep knowledge about government reporting, drawing on academic institutions and municipal securities analysts with decades of experience. We reviewed hundreds of ACFRs with a focus on general purpose governments, recognizing that to tackle the various special districts would require more than a volunteer group was able to take on, initially.

Second, we drew upon the XBRL data modeling skills among our members with technical and XBRL expertise. We developed three releases of the ACFR Taxonomy and published each iteration of the taxonomy for public review. We gathered and incorporated feedback received on appropriate definitions,

references, needed line-item captions and headers. Comments came in from government entities, standards organizations, and municipal securities analysts.

The ACFR Taxonomy at this stage contained government-wide, governmental funds, and proprietary funds statements (seven in total) plus notes for pensions and OPEB.

In 2021, we partnered with the University of Michigan's Center for Local State and Urban Policy (CLOSUP). CLOSUP identified funding for continued development from the Michigan-based Mott Foundation and from the University of Michigan Innovation Center for Academic Innovation which allowed us to conduct significant enhancements to the taxonomy.

Continued work to build a viable taxonomy representing general purpose governments included these steps:

- Ran an algorithm against thousands of Michigan financial statements to identify common line-item captions and headers across government financial statements. The algorithm allowed us to identify the number of times that a financial statement caption, for example, Delinquent Taxes Receivable, Noncurrent, appears on the Proprietary Funds Statement of Net Position. This helped to ensure that we captured the most common items on government financials.

- Analyzed the Michigan Chart of Accounts (COA) to capture needed line items on specific statements. All accounts in the COA were incorporated and references specific to Michigan were included as well.

- Engaged a California-based public sector CPA to review and assist in refining line items, presentation (ordering), GASB references, definitions, and labels for each statement covered.

- Expanded the taxonomy to include notes for Capital Assets, and Long-Term Debt.

- Prepared numerous sample XBRL reports to represent different general purpose governments to "test" the taxonomy and ensure that it was robust enough to capture all possible facts reported on a general purpose financial statement; and that it adhered to the structure of the data standard such that an XBRL-formatted financial statement would generate machine-readable, structured data.

- Conducted another public review to solicit further feedback. CLOSUP engaged with numerous Michigan-based issuer groups, Michigan State Treasury, and various accounting firms to solicit more feedback. All input was then incorporated back into the final release of the taxonomy which was published in late 2022.

We have also incorporated into the ACFR Taxonomy, line items that satisfy the requirements of public community colleges,

such as scholarship allowance and tuition revenue. XBRL reports have been developed and tested for entities including: the College of DuPage, William Rainey Harper College, and Oakton Community College.

Throughout this process, XBRL reports have been developed and tested for over 40 general purpose and special district (community college) governments. The taxonomy follows a modular approach with a base GASB Accounting Standards Taxonomy, which can support state-specific reporting requirements, government-specific reporting requirements (for example, for special districts), and also has a methodology to allow the reporting of government-specific line items that may be unique to that entity.

XBRL US is currently seeking additional funding to extend the taxonomy to other types of special purpose governments.

End notes

[1] H.R.7776, James M. Inhofe National Defense Authorization Act for Fiscal Year 2023 which includes TITLE LVIII, the Financial Data Transparency Act (FDTA): https://www.congress.gov/bill/117th-congress/house-bill/7776/text

[2] Warner & Crapo Introduce Legislation to Boost Transparency Around Financial Data, May 25, 2022: https://www.warner.senate.gov/public/index.cfm/2022/5/warner-crapo-introduce-legislation-to-boost-transparency-around-financial-data

[3] Office of Financial Research Interagency Data Inventory: https://www.financialresearch.gov/data/interagency-data-inventory/

[4] List of FDIC-Supervised Banks Filing under the Securities Exchange Act: https://www.fdic.gov/bank/individual/part335/

[5] Fragmented Securities Regulation, Information-Processing Costs, and Insider Trading: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3416204

[6] European Commission, Mergers: Commission approves acquisition of IHS Market by S&P Global, subject to conditions, October 22, 2021: https://ec.europa.eu/commission/presscorner/detail/en/ip_21_5461

[7] United States District Court Southern District of New York: https://storage.courtlistener.com/recap/gov.uscourts.nysd.576062/gov.uscourts.nysd.576062.70. 0.pdf

[8] XBRL Project Directory: https://www.xbrl.org/the-standard/why/xbrl-project-directory/

[9] XBRL International Open Information Model (OIM): https://specifications.xbrl.org/spec-group-index-open-information-model.html