DATA COALITION
An Initiative of the Data Foundation

Data Coalition
An Initiative of the Data Foundation
1003 K Street NW, Suite 200
Washington, D.C. 20001

@DataCoalition
info@datacoalition.org
datacoalition.org

Senator Sherrod Brown
Chair
Senate Banking Committee
Washington, D.C.

Senator Patrick Toomey
Ranking Member
Senate Banking Committee
Washington, D.C.

March 17th, 2021

Re: Proposal to Foster Economic Growth and Capital Formation

Dear Chairman Brown and Ranking Member Toomey,

Data Coalition proposes that the federal government modernize its financial regulatory data by enacting the Financial Transparency Act (FTA), which contains common-sense data reforms that benefit the capital markets, investors, and regulators alike. This legislation will enable policymakers and the American public to have access to reliable information about financial markets, building on other recent data reforms advanced in Congress like the 2014 Digital Accountability and Transparency Act (P.L. 113-101) and the Foundations for Evidence-Based Policymaking Act (P.L. 115-435).

In order to function efficiently, capital markets and investors require reliable and timely information. But most U.S. financial regulators do not use data standards to organize the information they collect from regulated entities. Regulators use inconsistent identifier codes for entities, instruments, and transactions. This means U.S. financial data have costly inefficiencies, resulting in duplicative and overlapping reporting requirements. This system creates inadequate data for capital markets, regulators, and retail investors and means higher costs to financial institutions and regulatory agencies alike.

The FTA addresses the longstanding data deficiencies in regulatory reporting by requiring the eight financial regulatory member-agencies of the U.S. Financial Stability Oversight Council to adopt and apply uniform standards for information already collected from regulated entities, including the adoption of a common, non-proprietary legal entity identifier. As a consequence, the data standards will enable better information processing, software-enabled filing preparation, and data reconciliation. These features collectively are the basis for retail investors, regulators, and the market having better information for selecting investment opportunities and understanding risks.

Better standardization of regulatory reporting requirements across the agencies would significantly improve the ability of the U.S. public sector to understand and identify the buildup of risk across financial products, institutions, and processes. Having good quality, standardized data is an important steppingstone to reaping the benefits of the ongoing digitization of financial assets, digitization of markets, and growing use of new, cutting-edge technologies, such as artificial intelligence. Reducing duplication, streamlining reporting, and using data standards would lead to efficiency, saving time, and reducing costs that firms and regulators otherwise expend manually collecting, reconciling, and consolidating data. The nonproprietary data standards in FTA will provide the basis for retail investors to have information that allows them to better select investment opportunities.

In addition to improving the data capital markets and investors rely on, the FTA can reduce compliance overhead for financial institutions. Streamlining regulatory reporting frees up valuable time and energy that can support private sector innovation and growth. Introducing the nonproprietary legal entity identifier decreases errors, improves the ability to draw comparisons across companies, and improves effective oversight. The Federal Reserve has extolled the benefits of applying such an identifier for these purposes.[1] If such an entity identifier had been in place decades ago, regulators may have been able to act more quickly in responding to the 2008 financial crisis. For example, better information would have been known in advance about the Lehman Brothers' practices and the risk to the financial system. With this information, financial firms and regulators would have been more able to more accurately and quickly assess the extent of damage and taken less radical action.

In the 116th Congress, the FTA had bipartisan support in the House of Representatives, co-sponsored by Representatives Carolyn Maloney (D-NY) and Patrick McHenry (R-NC). Adopting the FTA will improve financial data in the U.S. for the benefit of private markets, investors, and even regulators.

Attached is the legislative language, a summary of key provisions, as well as background materials. We strongly encourage the Senate Banking Committee to consider advancing sensible policies to align appropriate data standards, improve data quality, and reduce burden for regulatory reporting; the concepts in the Financial Transparency Act would go a long way toward achieving these goals.

Please feel free to contact me at corinna.turbes@datafoundation with any questions or concerns you may have.

Sincerely,

Corinna Turbes
Policy Director, Data Foundation

[1] Bottega, J., & Powell, L. (2011). *Creating a Linchpin for Financial Data: Toward a Universal Legal Entity Identifier* (Rep.). Washington, DC: Federal Reserve Board.

ATTACHMENTS:

Legislative Language

- Summary of Key Provisions
- H.R. 4476 Financial Transparency Act

Other Background Materials

- How Data Will Determine the Future of RegTech, by Donnelley Financial Solutions (DFIN), February 26, 2018
- Understanding Machine-Readability in Modern Data Policy, by Dean Ritz, Data Foundation and Senior Director of Structured Data Initiatives, Workiva, July 2020
- Standard Business Reporting: Open Data to Cut Compliance Costs, by Hudson Hollister et al, Data Foundation, March 2017
- Creating a Linchpin for Financial Data: Toward a Universal Legal Entity Identifier, John Bottega and Linda Powell, Federal Reserve Board, 2011

DATA COALITION

H.R. 4476: FINANCIAL TRANSPARENCY ACT (FTA) OF 2019

As Introduced in the House of Representatives on 9/24/2019

The proposed Financial Transparency Act directs major U.S. financial regulatory agencies to adopt consistent data fields and formats for information collected from industry. The FTA establishes a framework to improve the efficiency of regulatory reporting in coming years, reducing compliance overhead and the level of effort required for submitting financial reports. The required data standards will enable better information processing, software-enabled filing preparation, and data reconciliation. These features collectively are the basis for retail investors, regulators, and the market having better information for selecting investment opportunities and understanding risks.

* * * * * *

SUMMARY OF KEY PROVISIONS

Treasury and FSOC Member Agency Requirements

The FTA defines "data standard" and requires the Treasury Secretary to develop said "data standard." The Treasury Secretary can delegate the work to the appropriate Treasury entity, which would likely be the Financial Stability Oversight Council and the Office of Financial Research (OFR). *(Title I - Section 101)*

Common Identifiers: Transaction, Financial Instrument, Legal Entity

The FTA defines and requires the establishment of "common identifiers" for information reported to covered regulatory agencies, which could include transactions and financial products/instruments. It specifically requires the adoption of common, non-proprietary legal entity identifier for regulated organizations. The entity identifier would have to be made available under an "open license" which in existing law (per P.L. 115-435, Title II) means "a legal guarantee that a data asset is made available - at no cost to the public; and with no restrictions on copying, publishing, distributing, transmitting, citing, or adapting such asset." *(Title I - Section 101; Title II - Section 201; Title III - Section 301; Title IV - Section 401; Title V - Section 501; Title IV - Section 601; Title VII - Section 701; Title VIII - Section 801; Title IX - Section 901)*

Searchable, Machine-Readable, Open Data Standards

The FTA includes a set of required characteristics which builds upon industry and technology best practices, accounts for lessons learned from existing federal regulatory standard setting, and incorporates relevant federal policy and international standards definitions. The data standards require that data be rendered fully searchable, which is facilitated by the requirement to be "machine-readable" as now defined in federal law as meaning "data in a format that can be easily processed by a computer without human intervention while ensuring no semantic meaning is lost" (per P.L. 115-435). The data will be made available under "open license" format which will reduce barriers for industry, academia, and others to incorporate or reuse the data standards and information definitions into systems and processes. This requirement will also facilitate competition among multiple vendors for creation, data collection, and analysis tools, which reduce long-term costs. *(Title I - Section 101 & 102; Title II - Section 201-204; Title III - Section 301 & 302; Title IV - Section 401; Title V - Section 501; Title VI - Section 601 & 602; Title VII - Section 701 & 702; Title VIII - Section 801 & 802; Title IX - Section 901 & 902)*

Schemas and Modern Best Practices for Data Standards

The FTA requires definitional schemas as part of the data standards. Schemas and the accompanying metadata will need to be documented in machine-readable taxonomy and/or ontology models that clearly define the semantic meaning of the data. The schemas must also be mapped in machine-readable formats (e.g., XML, JSON, CSV) in order to facilitate large-scale, transactional-level information transfers. *(Title I - Section 101; Title II - Section 201; Title III - Section 301; Title IV - Section 401; Title V - Section 501; Title IV - Section 601; Title VII - Section 701; Title VIII - Section 801; Title IX - Section 901)*

Linking Data Standards to Regulatory Requirements

The FTA creates a reverse index from the collected regulatory data to the underlying regulatory information collection requirements. This will enable regulators and investors to query reported data based on regulatory collection requirements and improve reliability of data aggregation. *(Title I - Section 101; Title II - Section 201 & 205; Title III - Section 301; Title IV - Section 401; Title V - Section 501; Title IV - Section 601; Title VII - Section 701; Title VIII - Section 801; Title IX - Section 901)*

Leverage Existing Data Standards

The data standards required by FTA leverage existing, industry-accepted data formats and definitional standards. The standards connect with existing accounting standards to allow regulated entities to leverage expertise and processes established by the accounting, audit, legal, and regulatory compliance workforce. *(Title I - Section 101; Title II - Section 201; Title III - Section 301; Title IV - Section 401; Title V - Section 501; Title IV - Section 601; Title VII - Section 701; Title VIII - Section 801; Title IX - Section 901)*

Aligning with Government-Wide Policy for Open Data Publication

The FTA requires each covered agency to publish collected information, as appropriate, as "open data." The FTA reiterates the requirement for disclosable public data assets to be made available as "open Government Data asset" (per P.L. 115-435). This assures the data assets published under the regulatory authorities of the FTA's covered agencies are presented in a manner consistent with existing government-wide data policy ("machine-readable," "open license," and appropriate "metadata"). The FTA directs that these data sets are to be made available at the regulatory agencies for bulk downloaded in a "human readable format." *(Title I - Section 101 & 102; Title II - Section 201-204; Title III - Section 301 & 302; Title IV - Section 401; Title V - Section 501; Title VI - Section 601 & 602; Title VII - Section 701 & 702; Title VIII - Section 801 & 802; Title IX - Section 901 & 902)*

Implementation Timeframe

The U.S Treasury has two years from the date of enactment to develop and publish the required data standards. Covered agencies have three years to implement the data standards into their respective regulatory compliance reporting. *(Title I - Section 103; Title II - Section 201, 203 & 204; Title III - Section 303; Title IV - Section 402; Title V - Section 502; Title VI - Section 603; Title VII - Section 703; Title VIII - Section 803; Title IX - Section 903)*

Protections: Burden and Sensitive Information

The FTA establishes a "Scaling of Regulatory Requirements" in order to reduce the burden on smaller regulated entities. The FTA does not authorize any new regulatory information collections and does not authorize the publication of any information that is not already required to be published by existing law. The bill maintains exemptions under the Freedom of Information Act and existing protections in the Privacy Act. *(Title 1 - Section 104 & 105; Title II - Section 201, 203, 204 & 206; Title III - Section 303 & 304; Title IV - 402 & 403; Title V - Section 502 & 503; Title VI - Section 603 & 604; Title VII - Section 703 & 704; Title VIII - Section 803 & 804; Title IX - Section 904)*

DATA COALITION

H.R. 4476: FINANCIAL TRANSPARENCY ACT (FTA) OF 2019

As Introduced in the House of Representatives on 9/24/2019

The proposed Financial Transparency Act directs major U.S. financial regulatory agencies to adopt consistent data fields and formats for information collected from industry. The FTA establishes a framework to improve the efficiency of regulatory reporting in coming years, reducing compliance overhead and the level of effort required for submitting financial reports. The required data standards will enable better information processing, software-enabled filing preparation, and data reconciliation. These features collectively are the basis for retail investors, regulators, and the market having better information for selecting investment opportunities and understanding risks.

* * * * * *

SUMMARY OF KEY PROVISIONS

Treasury and FSOC Member Agency Requirements

The FTA defines "data standard" and requires the Treasury Secretary to develop said "data standard." The Treasury Secretary can delegate the work to the appropriate Treasury entity, which would likely be the Financial Stability Oversight Council and the Office of Financial Research (OFR). *(Title I - Section 101)*

Common Identifiers: Transaction, Financial Instrument, Legal Entity

The FTA defines and requires the establishment of "common identifiers" for information reported to covered regulatory agencies, which could include transactions and financial products/instruments. It specifically requires the adoption of common, non-proprietary legal entity identifier for regulated organizations. The entity identifier would have to be made available under an "open license" which in existing law (per P.L. 115-435, Title II) means "a legal guarantee that a data asset is made available - at no cost to the public; and with no restrictions on copying, publishing, distributing, transmitting, citing, or adapting such asset." *(Title I - Section 101; Title II - Section 201; Title III - Section 301; Title IV - Section 401; Title V - Section 501; Title IV - Section 601; Title VII - Section 701; Title VIII - Section 801; Title IX - Section 901)*

Searchable, Machine-Readable, Open Data Standards

The FTA includes a set of required characteristics which builds upon industry and technology best practices, accounts for lessons learned from existing federal regulatory standard setting, and incorporates relevant federal policy and international standards definitions. The data standards require that data be rendered fully searchable, which is facilitated by the requirement to be "machine-readable" as now defined in federal law as meaning "data in a format that can be easily processed by a computer without human intervention while ensuring no semantic meaning is lost" (per P.L. 115-435). The data will be made available under "open license" format which will reduce barriers for industry, academia, and others to incorporate or reuse the data standards and information definitions into systems and processes. This requirement will also facilitate competition among multiple vendors for creation, data collection, and analysis tools, which reduce long-term costs. *(Title I - Section 101 & 102; Title II - Section 201-204; Title III - Section 301 & 302; Title IV - Section 401; Title V - Section 501; Title VI - Section 601 & 602; Title VII - Section 701 & 702; Title VIII - Section 801 & 802; Title IX - Section 901 & 902)*

Schemas and Modern Best Practices for Data Standards

The FTA requires definitional schemas as part of the data standards. Schemas and the accompanying metadata will need to be documented in machine-readable taxonomy and/or ontology models that clearly define the semantic meaning of the data. The schemas must also be mapped in machine-readable formats (e.g., XML, JSON, CSV) in order to facilitate large-scale, transactional-level information transfers. *(Title I - Section 101; Title II - Section 201; Title III - Section 301; Title IV - Section 401; Title V - Section 501; Title IV - Section 601; Title VII - Section 701; Title VIII - Section 801; Title IX - Section 901)*

Linking Data Standards to Regulatory Requirements

The FTA creates a reverse index from the collected regulatory data to the underlying regulatory information collection requirements. This will enable regulators and investors to query reported data based on regulatory collection requirements and improve reliability of data aggregation. *(Title I - Section 101; Title II - Section 201 & 205; Title III - Section 301; Title IV - Section 401; Title V - Section 501; Title IV - Section 601; Title VII - Section 701; Title VIII - Section 801; Title IX - Section 901)*

Leverage Existing Data Standards

The data standards required by FTA leverage existing, industry-accepted data formats and definitional standards. The standards connect with existing accounting standards to allow regulated entities to leverage expertise and processes established by the accounting, audit, legal, and regulatory compliance workforce. *(Title I - Section 101; Title II - Section 201; Title III - Section 301; Title IV - Section 401; Title V - Section 501; Title IV - Section 601; Title VII - Section 701; Title VIII - Section 801; Title IX - Section 901)*

Aligning with Government-Wide Policy for Open Data Publication

The FTA requires each covered agency to publish collected information, as appropriate, as "open data." The FTA reiterates the requirement for disclosable public data assets to be made available as "open Government Data asset" (per P.L. 115-435). This assures the data assets published under the regulatory authorities of the FTA's covered agencies are presented in a manner consistent with existing government-wide data policy ("machine-readable," "open license," and appropriate "metadata"). The FTA directs that these data sets are to be made available at the regulatory agencies for bulk downloaded in a "human readable format." *(Title I - Section 101 & 102; Title II - Section 201-204; Title III - Section 301 & 302; Title IV - Section 401; Title V - Section 501; Title VI - Section 601 & 602; Title VII - Section 701 & 702; Title VIII - Section 801 & 802; Title IX - Section 901 & 902)*

Implementation Timeframe

The U.S Treasury has two years from the date of enactment to develop and publish the required data standards. Covered agencies have three years to implement the data standards into their respective regulatory compliance reporting. *(Title I - Section 103; Title II - Section 201, 203 & 204; Title III - Section 303; Title IV - Section 402; Title V - Section 502; Title VI - Section 603; Title VII - Section 703; Title VIII - Section 803; Title IX - Section 903)*

Protections: Burden and Sensitive Information

The FTA establishes a "Scaling of Regulatory Requirements" in order to reduce the burden on smaller regulated entities. The FTA does not authorize any new regulatory information collections and does not authorize the publication of any information that is not already required to be published by existing law. The bill maintains exemptions under the Freedom of Information Act and existing protections in the Privacy Act. *(Title 1 - Section 104 & 105; Title II - Section 201, 203, 204 & 206; Title III - Section 303 & 304; Title IV - 402 & 403; Title V - Section 502 & 503; Title VI - Section 603 & 604; Title VII - Section 703 & 704; Title VIII - Section 803 & 804; Title IX - Section 904)*

I

116TH CONGRESS
1ST SESSION

H. R. 4476

To amend securities, commodities, and banking laws to make the information reported to financial regulatory agencies electronically searchable, to further enable the development of RegTech 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.

IN THE HOUSE OF REPRESENTATIVES

SEPTEMBER 24, 2019

Mrs. CAROLYN B. MALONEY of New York (for herself and Mr. MCHENRY) introduced the following bill; which was referred to the Committee on Financial Services, and in addition to the Committee on Agriculture, for a period to be subsequently determined by the Speaker, in each case for consideration of such provisions as fall within the jurisdiction of the committee concerned

---

A BILL

To amend securities, commodities, and banking laws to make the information reported to financial regulatory agencies electronically searchable, to further enable the development of RegTech 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.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE; TABLE OF CONTENTS.

(a) SHORT TITLE.—This Act may be cited as the ''Financial Transparency Act of 2019''.

(b) TABLE OF CONTENTS.—The table of contents for this Act is as follows:

TITLE I—DEPARTMENT OF THE TREASURY

SEC. 101. DATA STANDARDS.

(a) IN GENERAL.—Subtitle A of title I of the Financial Stability Act of 2010 (12 U.S.C. 5311 et seq.) is amended by adding at the end the following:

"SEC. 124. DATA STANDARDS.

''(a) IN GENERAL.—The Secretary of the Treasury shall, by rule, promulgate data standards, meaning a

standard that specifies rules by which data is described and recorded, for the information reported to member agencies by financial entities under the jurisdiction of the member agency and the data collected from member agencies on behalf of the Council.

''(b) STANDARDIZATION.—Member agencies, in consultation with the Secretary of the Treasury, shall implement regulations promulgated by the Secretary of the Treasury under subsection (a) to standardize data reported to member agencies or collected on behalf of the Council, as described under subsection (a). If a member agency fails to implement such regulations prior to the expiration of the 3-year period following the date of publication of final regulations, the Secretary of the Treasury, in consultation with the Chairperson, may implement such regulations with respect to the financial entities under the jurisdiction of the member agency.

''(c) DATA STANDARDS.—

''(1) COMMON IDENTIFIERS.—The data standards promulgated under subsection (a) shall include common identifiers for information reported to member agencies or collected on behalf of the Council, including a common nonproprietary legal entity identifier that is available under an open license (as defined under section 3502 of title 44, United States

Code) for all entities required to report to member agencies.

''(2) DATA STANDARD.—The data standards promulgated under subsection (a) shall, to the extent practicable—

''(A) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(B) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

''(C) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

''(D) be nonproprietary or made available under an open license;

''(E) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(F) use, be consistent with, and implement applicable accounting and reporting principles.

''(3) CONSULTATION.—In promulgating data standards under subsection (a), the Secretary of the Treasury shall consult with other Federal departments and agencies and multi-agency initiatives responsible for Federal data standards.

''(4) INTEROPERABILITY OF DATA.—In promulgating data standards under subsection (a), the Secretary of the Treasury shall seek to promote interoperability of financial regulatory data across members of the Council.''.

(b) CLERICAL AMENDMENT.—The table of contents under section 1(b) of the Dodd-Frank Wall Street Reform and Consumer Protection Act is amended by inserting after the item relating to section 123 the following:

''Sec. 124. Data standards.''.

SEC. 102. OPEN DATA PUBLICATION BY THE DEPARTMENT OF THE TREASURY.

Section 124 of the Financial Stability Act of 2010, as added by section 101, is amended by adding at the end the following:

''(d) OPEN DATA PUBLICATION.—All public information published by the Secretary of the Treasury under this subtitle shall be made available as an open Government data asset (as defined under section 3502 of title 44, United States Code), freely available for download in bulk, and rendered in a human-readable format and accessible via application programming interface where appropriate.''.

SEC. 103. RULEMAKING.

Not later than the end of the 2-year period beginning on the date of the enactment of this Act, the Secretary of the Treasury shall issue the regulations required under the amendments made by this title. The Secretary may delegate the functions required under the amendments made by this title to an appropriate office within the Department of the Treasury.

SEC. 104. CLASSIFIED AND PROTECTED INFORMATION.

Nothing in this title or the amendments made by this title shall require the disclosure to the public of—

(1) information that would be exempt from disclosure under section 552 of title 5, United States Code (commonly known as the ''Freedom of Information Act''); or

(2) information protected under section 552a of title 5, United States Code (commonly known as the

''Privacy Act of 1974''), or section 6103 of the Internal Revenue Code of 1986.

SEC. 105. NO NEW DISCLOSURE REQUIREMENTS.

Nothing in this title or the amendments made by this title shall be construed to require the Secretary of the Treasury to collect additional information under the statutes amended by this title, beyond information that was collected under such statutes before the date of the enactment of this Act.

SEC. 106. REPORT.

Not later than 1 year after the end of the 2-year period described in section 103, the Comptroller General of the United States shall submit to Congress a report on the feasibility, costs, and potential benefits of building upon the taxonomy established by this Act to arrive at a Federal Government-wide regulatory compliance standardization mechanism similar to Standard Business Reporting.

SEC. 107. CONFORMING AMENDMENTS TO THE OFFICE OF FINANCIAL RESEARCH.

Section 153 of the Financial Stability Act of 2010 (12 U.S.C. 5343) is amended—

(1) in subsection (a)—

(A) by striking paragraph (2); and

(B) by redesignating paragraphs (3) through (7) as paragraphs (2) through (6), respectively; and

(2) by amending subsection (c) to read as follows:

''(c) RULEMAKING AUTHORITY.—The Office, in consultation with the Chairperson, shall issue rules, regulations, and orders only to the extent necessary to carry out the purposes and duties described in paragraphs (1) and (6) of subsection (a).''.

TITLE II—SECURITIES AND EXCHANGE COMMISSION

SEC. 201. DATA STANDARDS REQUIREMENTS FOR THE SECURITIES AND EXCHANGE COMMISSION.

(a) DATA STANDARDS FOR INVESTMENT ADVISERS' REPORTS UNDER THE INVESTMENT ADVISERS ACT OF 1940.—Section 204 of the Investment Advisers Act of 1940 (15 U.S.C. 80b–4) is amended—

(1) by redesignating the second subsection (d) (relating to Records of Persons With Custody of Use) as subsection (e); and

(2) by adding at the end the following:

''(f) DATA STANDARDS FOR REPORTS FILED UNDER THIS SECTION.—

''(1) REQUIREMENT.—The Commission shall, by rule, adopt data standards for all reports filed by investment advisers with the Commission under this section.

''(2) CHARACTERISTICS.—The data standards required by paragraph (1) shall, to the extent practicable—

''(A) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(B) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

''(C) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

''(D) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

''(E) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(F) use, be consistent with, and implement applicable accounting and reporting principles.

''(3) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this subsection, the Commission shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.''.

(b) DATA STANDARDS FOR REGISTRATION STATEMENTS AND REPORTS UNDER THE INVESTMENT COMPANY ACT OF 1940.—The Investment Company Act of 1940 (15 U.S.C. 80a–1 et seq.) is amended—

(1) in section 8, by adding at the end the following:

''(g) DATA STANDARDS FOR REGISTRATION STATEMENTS.—

''(1) REQUIREMENT.—The Commission shall, by rule, adopt data standards for all registration statements required to be filed with the Commission

under this section, except that the Commission may exempt exhibits, signatures, and certifications from such data standards.

''(2) CHARACTERISTICS.—The data standards required by paragraph (1) shall, to the extent practicable—

''(A) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(B) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

''(C) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

''(D) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

''(E) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(F) use, be consistent with, and implement applicable accounting and reporting principles.

''(3) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this subsection, the Commission shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.''; and

(2) in section 30, by adding at the end the following:

''(k) DATA STANDARDS FOR REPORTS.—

''(1) REQUIREMENT.—The Commission shall, by rule, adopt data standards for all reports required to be filed with the Commission under this section, except that the Commission may exempt exhibits, signatures, and certifications from such data standards.

''(2) CHARACTERISTICS.—The data standards required by paragraph (1) shall, to the extent practicable—

"(A) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

"(B) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

"(C) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

"(D) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

"(E) incorporate standards developed and maintained by voluntary consensus standards bodies; and

"(F) use, be consistent with, and implement applicable accounting and reporting principles.

''(3) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this subsection, the Commission shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.''.

(c) DATA STANDARDS FOR INFORMATION REQUIRED TO BE SUBMITTED OR PUBLISHED BY NATIONALLY RECOGNIZED STATISTICAL RATING ORGANIZATIONS.—Section 15E of the Securities Exchange Act of 1934 (15 U.S.C. 78o–7) is amended by adding at the end the following:

''(w) DATA STANDARDS FOR INFORMATION REQUIRED TO BE SUBMITTED OR PUBLISHED UNDER THIS SECTION.—

''(1) REQUIREMENT.—The Commission shall, by rule, adopt data standards for all information required to be submitted or published by a nationally recognized statistical rating organization under this section.

''(2) CHARACTERISTICS.—The data standards required by paragraph (1) shall, to the extent practicable—

''(A) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

"(B) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

"(C) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

"(D) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

"(E) incorporate standards developed and maintained by voluntary consensus standards bodies; and

"(F) use, be consistent with, and implement applicable accounting and reporting principles.

"(3) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this subsection, the Commission shall incorporate all applica-

ble data standards promulgated by the Secretary of the Treasury.''.

(d) DATA STANDARDS FOR ASSET-BACKED SECURITIES DISCLOSURES.—Section 7(c) of the Securities Act of 1933 (15 U.S.C. 77g(c)) is amended by adding at the end the following:

''(3) DATA STANDARDS FOR ASSET-BACKED SECURITIES DISCLOSURES.—

''(A) REQUIREMENT.—The Commission shall, by rule, adopt data standards for all disclosures required under this subsection.

''(B) CHARACTERISTICS.—The data standards required by subparagraph (A) shall, to the extent practicable—

''(i) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(ii) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying

regulatory information collection requirements;

''(iii) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

''(iv) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

''(v) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(vi) use, be consistent with, and implement applicable accounting and reporting principles.

''(C) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this paragraph, the Commission shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.''.

(e) DATA STANDARDS FOR CORPORATE DISCLOSURES UNDER THE SECURITIES ACT OF 1933.—Section

7 of the Securities Act of 1933 (15 U.S.C. 77g) is amended by adding at the end the following:

''(e) DATA STANDARDS.—

''(1) REQUIREMENT.—The Commission shall, by rule, adopt data standards for all registration statements and for all prospectuses included in registration statements required to be filed with the Commission under this title, except that the Commission may exempt exhibits, signatures, and certifications from such data standards.

''(2) CHARACTERISTICS.—The data standards required by paragraph (1) shall, to the extent practicable—

''(A) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(B) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

''(C) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

''(D) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

''(E) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(F) use, be consistent with, and implement applicable accounting and reporting principles.

''(3) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this subsection, the Commission shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.''.

(f) DATA STANDARDS FOR PERIODIC AND CURRENT CORPORATE DISCLOSURES UNDER THE SECURITIES EXCHANGE ACT OF 1934.—Section 13 of the Securities Exchange Act of 1934 (15 U.S.C. 78m) is amended by adding at the end the following:

''(s) DATA STANDARDS.—

''(1) REQUIREMENT.—The Commission shall, by rule, adopt data standards for all information contained in periodic and current reports required to be filed or furnished under this section or under section 15(d), except that the Commission may exempt exhibits, signatures, and certifications from such data standards.

''(2) CHARACTERISTICS.—The data standards required by paragraph (1) shall, to the extent practicable—

''(A) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(B) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

''(C) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be

consistently identified as such in associated machine-readable metadata;

''(D) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

''(E) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(F) use, be consistent with, and implement applicable accounting and reporting principles.

''(3) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this subsection, the Commission shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.''.

(g) DATA STANDARDS FOR CORPORATE PROXY AND CONSENT SOLICITATION MATERIALS UNDER THE SECURITIES EXCHANGE ACT OF 1934.—Section 14 of the Securities Exchange Act of 1934 (15 U.S.C. 78n) is amended by adding at the end the following:

''(k) DATA STANDARDS FOR PROXY AND CONSENT SOLICITATION MATERIALS.—

''(1) REQUIREMENT.—The Commission shall, by rule, adopt data standards for all information

contained in any proxy or consent solicitation material prepared by an issuer for an annual meeting of the shareholders of the issuer, except that the Commission may exempt exhibits, signatures, and certifications from such data standards.

''(2) CHARACTERISTICS.—The data standards required by paragraph (1) shall, to the extent practicable—

''(A) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(B) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

''(C) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

''(D) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

''(E) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(F) use, be consistent with, and implement applicable accounting and reporting principles.

''(3) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this subsection, the Commission shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.''.

(h) DATA STANDARDS FOR SECURITY-BASED SWAP REPORTING.—Section 15F of the Securities Exchange Act of 1934 (15 U.S.C. 78o–10) is amended by adding at the end the following:

''(m) DATA STANDARDS FOR SECURITY-BASED SWAP REPORTING.—

''(1) REQUIREMENT.—The Commission shall, by rule, adopt data standards for all reports related to security-based swaps that are required under this Act.

''(2) CHARACTERISTICS.—The data standards required by paragraph (1) shall, to the extent practicable—

''(A) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(B) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

''(C) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

''(D) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

''(E) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(F) use, be consistent with, and implement applicable accounting and reporting principles.

''(3) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this subsection, the Commission shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.''.

(i) RULEMAKING.—

(1) IN GENERAL.—Not later than the end of the 2-year period beginning on the date of the enactment of this Act, the Securities and Exchange Commission shall issue the regulations required under the amendments made by this section.

(2) SCALING OF REGULATORY REQUIREMENTS.—In issuing the regulations required under the amendments made by this section, the Securities and Exchange Commission may scale data reporting requirements in order to reduce any unjustified burden on emerging growth companies, lending institutions, accelerated filers, smaller reporting companies, and other smaller issuers, as determined by the study required under section 205(c), while still providing searchable information to investors.

(3) MINIMIZING DISRUPTION.—In issuing the regulations required under the amendments made by this section, the Securities and Exchange Commission shall seek to minimize disruptive changes to the persons affected by such regulations.

SEC. 202. OPEN DATA PUBLICATION BY THE SECURITIES AND EXCHANGE COMMISSION.

Section 4 of the Securities Exchange Act of 1934 (15 U.S.C. 78d) is amended by adding at the end the following:

''(j) OPEN DATA PUBLICATION.—All public information published by the Commission under the securities laws and the Dodd-Frank Wall Street Reform and Consumer Protection Act shall be made available as an open Government data asset (as defined under section 3502 of title 44, United States Code), freely available for download in bulk and rendered in a human-readable format and accessible via application programming interface where appropriate.''.

SEC. 203. DATA TRANSPARENCY AT THE MUNICIPAL SECURITIES RULEMAKING BOARD.

(a) IN GENERAL.—Section 15B(b) of the Securities Exchange Act of 1934 (15 U.S.C. 78o–4(b)) is amended by adding at the end the following:

''(8) DATA STANDARDS.—

''(A) REQUIREMENT.—If the Board establishes information systems under paragraph (3), the Board shall adopt data standards for information submitted via such systems.

''(B) CHARACTERISTICS.—The data standards required by subparagraph (A) shall, to the extent practicable—

''(i) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(ii) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

''(iii) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

''(iv) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

''(v) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(vi) use, be consistent with, and implement applicable accounting and reporting principles.

''(C) INCORPORATION OF STANDARDS.—In adopting data standards under this paragraph, the Board shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.''.

(b) RULEMAKING.—

(1) IN GENERAL.—Not later than the end of the 2-year period beginning on the date of the enactment of this Act, the Municipal Securities Rulemaking Board shall issue the regulations required under the amendments made by this section.

(2) SCALING OF REGULATORY REQUIREMENTS.—In issuing the regulations required under the amendments made by this section, the Municipal Securities Rulemaking Board may scale data reporting requirements in order to reduce any unjustified burden on smaller regulated entities.

(3) MINIMIZING DISRUPTION.—In issuing the regulations required under the amendments made by this section, the Municipal Securities Rulemaking Board shall seek to minimize disruptive changes to the persons affected by such regulations.

SEC. 204. DATA TRANSPARENCY AT NATIONAL SECURITIES ASSOCIATIONS.

(a) IN GENERAL.—Section 15A of the Securities Exchange Act of 1934 (15 U.S.C. 78o–3) is amended by adding at the end the following:

''(n) DATA STANDARDS.—

''(1) REQUIREMENT.—A national securities association registered pursuant to subsection (a) shall adopt data standards for all information that is regularly filed with or submitted to the association.

''(2) CHARACTERISTICS.—The data standards required by paragraph (1) shall, to the extent practicable—

''(A) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(B) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable

taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

"(C) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

"(D) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

"(E) incorporate standards developed and maintained by voluntary consensus standards bodies; and

"(F) use, be consistent with, and implement applicable accounting and reporting principles.

"(3) INCORPORATION OF STANDARDS.—In adopting data standards under this subsection, the association shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.".

(b) RULEMAKING.—

(1) IN GENERAL.—Not later than the end of the 2-year period beginning on the date of the enactment of this Act, a national securities association shall adopt the standards required under the amendments made by this section.

(2) SCALING OF REGULATORY REQUIREMENTS.—In adopting the standards required under the amendments made by this section, a national securities association may scale data reporting requirements in order to reduce any unjustified burden on smaller regulated entities.

(3) MINIMIZING DISRUPTION.—In adopting the standards required under the amendments made by this section, a national securities association shall seek to minimize disruptive changes to the persons affected by such standards.

SEC. 205. SHORTER-TERM BURDEN REDUCTION AND DISCLOSURE SIMPLIFICATION AT THE SECURITIES AND EXCHANGE COMMISSION; SUNSET.

(a) BETTER ENFORCEMENT OF THE QUALITY OF CORPORATE FINANCIAL DATA SUBMITTED TO THE SECURITIES AND EXCHANGE COMMISSION.—

(1) DATA QUALITY IMPROVEMENT PROGRAM.—Within six months after the date of the enactment of this Act, the Commission shall establish a pro-

gram to improve the quality of corporate financial data filed or furnished by issuers under the Securities Act of 1933, the Securities Exchange Act of 1934, and the Investment Company Act of 1940. The program shall include the following:

(A) The designation of an official in the Office of the Chairman responsible for the improvement of the quality of data filed with or furnished to the Commission by issuers.

(B) The issuance by the Division of Corporation Finance of comment letters requiring correction of errors in data filings and submissions, where necessary.

(2) GOALS.—In establishing the program under this section, the Commission shall seek to—

(A) improve the quality of data filed with or furnished to the Commission to a commercially acceptable level; and

(B) make data filed with or furnished to the Commission useful to investors.

(b) REPORT ON THE USE OF MACHINE-READABLE DATA FOR CORPORATE DISCLOSURES.—

(1) IN GENERAL.—Not later than six months after the date of the enactment of this Act, and every six months thereafter, the Commission shall

issue a report to the Committee on Financial Services of the House of Representatives and the Committee on Banking, Housing, and Urban Affairs of the Senate on the public and internal use of machine-readable data for corporate disclosures.

(2) CONTENT.—Each report required under paragraph (1) shall include—

(A) an identification of which corporate disclosures required under section 7 of the Securities Act of 1933, section 13 of the Securities Exchange Act of 1934, or section 14 of the Securities Exchange Act of 1934 are expressed as machine-readable data and which are not;

(B) an analysis of the costs and benefits of the use of machine-readable data in corporate disclosure to investors, markets, the Commission, and issuers;

(C) a summary of enforcement actions that result from the use or analysis of machine-readable data collected under section 7 of the Securities Act of 1933, section 13 of the Securities Exchange Act of 1934, or section 14 of the Securities Exchange Act of 1934; and

(D) an analysis of how the Commission is itself using the machine-readable data collected by the Commission.

(c) SUNSET.—On and after the end of the 7-year period beginning on the date of the enactment of this Act, this section shall have no force or effect.

SEC. 206. NO NEW DISCLOSURE REQUIREMENTS.

Nothing in this title or the amendments made by this title shall be construed to require the Securities and Exchange Commission, the Municipal Securities Rulemaking Board, or a national securities association to collect additional information under the statutes amended by this title, beyond information that was collected under such statutes before the date of the enactment of this Act.

TITLE III—FEDERAL DEPOSIT INSURANCE CORPORATION

SEC. 301. DATA STANDARDS REQUIREMENTS FOR THE FEDERAL DEPOSIT INSURANCE CORPORATION.

The Federal Deposit Insurance Act (12 U.S.C. 1811 et seq.) is amended by adding at the end the following:

"SEC. 52. DATA STANDARDS.

''(a) REQUIREMENT.—The Corporation shall, by rule, adopt data standards for all information that the Corporation receives from any depository institution or financial

company under this Act or under title II of the Dodd-Frank Wall Street Reform and Consumer Protection Act.

''(b) CHARACTERISTICS.—The data standards required by subsection (a) shall, to the extent practicable—

''(1) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(2) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

''(3) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

''(4) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

''(5) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(6) use, be consistent with, and implement applicable accounting and reporting principles.

''(c) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this section, the Corporation shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.

''(d) FINANCIAL COMPANY DEFINED.—For purposes of this section, the term 'financial company' has the meaning given that term under section 201(a) of the Dodd-Frank Wall Street Reform and Consumer Protection Act (12 U.S.C. 5381(a)).''.

SEC. 302. OPEN DATA PUBLICATION BY THE FEDERAL DEPOSIT INSURANCE CORPORATION.

The Federal Deposit Insurance Act (12 U.S.C. 1811 et seq.), as amended by section 301, is further amended by adding at the end the following:

''SEC. 53. OPEN DATA PUBLICATION.

''All public information published by the Corporation under this Act or under the Dodd-Frank Wall Street Reform and Consumer Protection Act shall be made available as an open Government data asset (as defined under section 3502 of title 44, United States Code), freely available for download in bulk and rendered in a human-readable format and accessible via application programming interface where appropriate.''.

SEC. 303. RULEMAKING.

(a) IN GENERAL.—Not later than the end of the 2-year period beginning on the date of the enactment of this Act, the Federal Deposit Insurance Corporation shall issue the regulations required under the amendments made by this title.

(b) SCALING OF REGULATORY REQUIREMENTS.—In issuing the regulations required under the amendments made by this title, the Federal Deposit Insurance Corporation may scale data reporting requirements in order to reduce any unjustified burden on smaller regulated entities.

(c) MINIMIZING DISRUPTION.—In issuing the regulations required under the amendments made by this title, the Federal Deposit Insurance Corporation shall seek to minimize disruptive changes to the persons affected by such regulations.

SEC. 304. NO NEW DISCLOSURE REQUIREMENTS.

Nothing in this title or the amendments made by this title shall be construed to require the Federal Deposit Insurance Corporation to collect additional information under the statutes amended by this title, beyond information that was collected under such statutes before the date of the enactment of this Act.

TITLE IV—OFFICE OF THE COMPTROLLER OF THE CURRENCY

SEC. 401. DATA STANDARDS AND OPEN DATA PUBLICATION REQUIREMENTS FOR THE OFFICE OF THE COMPTROLLER OF THE CURRENCY.

The Revised Statutes of the United States is amended by inserting after section 332 (12 U.S.C. 14) the following:

"SEC. 333. DATA STANDARDS; OPEN DATA PUBLICATION.

''(a) DATA STANDARDS.—

''(1) REQUIREMENT.—The Comptroller of the Currency shall, by rule, adopt data standards for all information that is regularly filed with or submitted to the Comptroller of the Currency by any entity with respect to which the Office of the Comptroller of the Currency is the appropriate Federal banking agency (as defined under section 3 of the Federal Deposit Insurance Act).

''(2) CHARACTERISTICS.—The data standards required by paragraph (1) shall, to the extent practicable—

''(A) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

"(B) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

"(C) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

"(D) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

"(E) incorporate standards developed and maintained by voluntary consensus standards bodies; and

"(F) use, be consistent with, and implement applicable accounting and reporting principles.

"(3) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this subsection, the Comptroller of the Currency shall incor-

porate all applicable data standards promulgated by the Secretary of the Treasury.

''(b) OPEN DATA PUBLICATION.—All public information published by the Comptroller of the Currency under title LXII or the Dodd-Frank Wall Street Reform and Consumer Protection Act shall be made available as an open Government data asset (as defined under section 3502 of title 44, United States Code), freely available for download in bulk and rendered in a human-readable format and accessible via application programming interface where appropriate.''.

SEC. 402. RULEMAKING.

(a) IN GENERAL.—Not later than the end of the 2-year period beginning on the date of the enactment of this Act, the Comptroller of the Currency shall issue the regulations required under the amendments made by this title.

(b) SCALING OF REGULATORY REQUIREMENTS.—In issuing the regulations required under the amendments made by this title, the Comptroller of the Currency may scale data reporting requirements in order to reduce any unjustified burden on smaller regulated entities.

(c) MINIMIZING DISRUPTION.—In issuing the regulations required under the amendments made by this title, the Comptroller of the Currency shall seek to minimize

disruptive changes to the persons affected by such regulations.

SEC. 403. NO NEW DISCLOSURE REQUIREMENTS.

Nothing in this title or the amendments made by this title shall be construed to require the Comptroller of the Currency to collect additional information under the statutes amended by this title, beyond information that was collected under such statutes before the date of the enactment of this Act.

TITLE V—BUREAU OF CONSUMER FINANCIAL PROTECTION

SEC. 501. DATA STANDARDS AND OPEN DATA PUBLICATION REQUIREMENTS FOR THE BUREAU OF CONSUMER FINANCIAL PROTECTION.

(a) IN GENERAL.—The Consumer Financial Protection Act of 2010 (12 U.S.C. 5481 et seq.) is amended by inserting after section 1018 the following:

"SEC. 1019. DATA STANDARDS.

"(a) REQUIREMENT.—The Bureau shall, by rule, adopt data standards for all information that is regularly filed with or submitted to the Bureau under this title.

"(b) CHARACTERISTICS.—The data standards required by subsection (a) shall, to the extent practicable—

''(1) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(2) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

''(3) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

''(4) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

''(5) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(6) use, be consistent with, and implement applicable accounting and reporting principles.

''(c) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this section, the Bureau

shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.

''SEC. 1020. OPEN DATA PUBLICATION.

''All public information published by the Bureau under this title shall be made available as an open Government data asset (as defined under section 3502 of title 44, United States Code), freely available for download in bulk and rendered in a human-readable format and accessible via application programming interface where appropriate.''.

(b) CLERICAL AMENDMENT.—The table of contents under section 1(b) of the Dodd-Frank Wall Street Reform and Consumer Protection Act is amended by inserting after the item relating to section 1018 the following:

''Sec. 1019. Data standards.
''Sec. 1020. Open data publication.''.

SEC. 502. RULEMAKING.

(a) IN GENERAL.—Not later than the end of the 2-year period beginning on the date of the enactment of this Act, the Bureau of Consumer Financial Protection shall issue the regulations required under the amendments made by this title.

(b) SCALING OF REGULATORY REQUIREMENTS.—In issuing the regulations required under the amendments made by this title, the Bureau of Consumer Financial Protection may scale data reporting requirements in order to

reduce any unjustified burden on smaller regulated entities.

(c) MINIMIZING DISRUPTION.—In issuing the regulations required under the amendments made by this title, the Bureau of Consumer Financial Protection shall seek to minimize disruptive changes to the persons affected by such regulations.

SEC. 503. NO NEW DISCLOSURE REQUIREMENTS.

Nothing in this title or the amendments made by this title shall be construed to require the Bureau of Consumer Financial Protection to collect additional information under the statutes amended by this title, beyond information that was collected under such statutes before the date of the enactment of this Act.

TITLE VI—FEDERAL RESERVE SYSTEM

SEC. 601. DATA STANDARDS REQUIREMENTS FOR THE BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM.

(a) DATA STANDARDS FOR INFORMATION FILED OR SUBMITTED BY NONBANK FINANCIAL COMPANIES.—Section 161(a) of the Financial Stability Act of 2010 (12 U.S.C. 5361(a)) is amended by adding at the end the following:

''(4) DATA STANDARDS FOR REPORTS UNDER THIS SUBSECTION.—

''(A) IN GENERAL.—The Board of Governors shall adopt data standards for all financial data that is regularly filed with or submitted to the Board of Governors by any nonbank financial company pursuant to this subsection.

''(B) CHARACTERISTICS.—The data standards required by this section shall, to the extent practicable—

''(i) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(ii) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

''(iii) assure that a data element or data asset that exists to satisfy an under-

lying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

''(iv) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

''(v) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(vi) use, be consistent with, and implement applicable accounting and reporting principles.

''(C) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this paragraph, the Board of Governors shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.''.

(b) DATA STANDARDS FOR INFORMATION FILED OR SUBMITTED BY SAVINGS AND LOAN HOLDING COMPANIES.—Section 10 of the Home Owners' Loan Act (12 U.S.C. 1467a) is amended by adding at the end the following:

''(u) DATA STANDARDS.—

''(1) REQUIREMENT.—The Board shall adopt data standards for all information that is regularly filed with or submitted to the Board by any savings and loan holding company, or subsidiary of a savings and loan holding company, other than a depository institution, under this section.

''(2) CHARACTERISTICS.—The data standards required by this subsection shall, to the extent practicable—

''(A) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(B) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

''(C) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

''(D) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

''(E) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(F) use, be consistent with, and implement applicable accounting and reporting principles.

''(3) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this section, the Board of Governors shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.''.

(c) DATA STANDARDS FOR INFORMATION FILED OR SUBMITTED BY BANK HOLDING COMPANIES.—Section 5 of the Bank Holding Company Act of 1956 (12 U.S.C. 1844) is amended by adding at the end the following:

''(h) DATA STANDARDS.—

''(1) REQUIREMENT.—The Board shall adopt data standards for all information that is regularly filed with or submitted to the Board by any bank holding company in a report under subsection (c).

''(2) CHARACTERISTICS.—The data standards required by this subsection shall, to the extent practicable—

''(A) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(B) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

''(C) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

''(D) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

''(E) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(F) use, be consistent with, and implement applicable accounting and reporting principles.

''(3) INCORPORATION OF STANDARDS.—In adopting data standards under this subsection, the Board shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.''.

(d) DATA STANDARDS FOR INFORMATION SUBMITTED BY FINANCIAL MARKET UTILITIES OR INSTITUTIONS UNDER THE PAYMENT, CLEARING, AND SETTLEMENT SUPERVISION ACT OF 2010.—Section 809 of the Payment, Clearing, and Settlement Supervision Act of 2010 (12 U.S.C. 5468) is amended by adding at the end the following:

''(h) DATA STANDARDS.—

''(1) REQUIREMENT.—The Board of Governors shall adopt data standards for all information that is regularly filed with or submitted to the Board by any financial market utility or financial institution under subsection (a) or (b).

''(2) CHARACTERISTICS.—The data standards required by this subsection shall, to the extent practicable—

"(A) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

"(B) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

"(C) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

"(D) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

"(E) incorporate standards developed and maintained by voluntary consensus standards bodies; and

"(F) use, be consistent with, and implement applicable accounting and reporting principles.

''(3) INCORPORATION OF STANDARDS.—In adopting data standards under this subsection, the Board of Governors shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.''.

SEC. 602. OPEN DATA PUBLICATION BY THE BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM.

The Federal Reserve Act (12 U.S.C. 226 et seq.) is amended by adding at the end the following:

''SEC. 32. OPEN DATA PUBLICATION BY THE BOARD OF GOVERNORS.

''All public information published by the Board of Governors under this Act, the Bank Holding Company Act of 1956, the Financial Stability Act of 2010, the Home Owners' Loan Act, the Payment, Clearing, and Settlement Supervision Act of 2010, or the Enhancing Financial Institution Safety and Soundness Act of 2010 shall be made available as an open Government data asset (as defined under section 3502 of title 44, United States Code), freely available for download in bulk and rendered in a human-readable format and accessible via application programming interface where appropriate.''.

SEC. 603. RULEMAKING.

(a) IN GENERAL.—Not later than the end of the 2-year period beginning on the date of the enactment of this Act, the Board of Governors of the Federal Reserve System shall issue the regulations required under the amendments made by this title.

(b) SCALING OF REGULATORY REQUIREMENTS.—In issuing the regulations required under the amendments made by this title, the Board of Governors of the Federal Reserve System may scale data reporting requirements in order to reduce any unjustified burden on smaller regulated entities.

(c) MINIMIZING DISRUPTION.—In issuing the regulations required under the amendments made by this title, the Board of Governors of the Federal Reserve System shall seek to minimize disruptive changes to the persons affected by such regulations.

SEC. 604. NO NEW DISCLOSURE REQUIREMENTS.

Nothing in this title or the amendments made by this title shall be construed to require the Board of Governors of the Federal Reserve System to collect additional information under the statutes amended by this title, beyond information that was collected under such statutes before the date of the enactment of this Act.

TITLE VII—COMMODITY FUTURES TRADING COMMISSION

SEC. 701. DATA STANDARDS.

The Commodity Exchange Act (7 U.S.C. 1 et seq.) is amended by adding at the end the following:

"SEC. 24. DATA STANDARDS.

''(a) REQUIREMENT.—The Commission shall, by rule, adopt data standards for all information that is regularly filed with or submitted to the Commission under this Act, all information that is required to be reported to a registered swap data repository under this Act, and all information that is required to be publicly disclosed by parties to a swap under this Act.

''(b) CHARACTERISTICS.—The data standards required by subsection (a) shall, to the extent practicable—

''(1) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(2) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic

meaning as defined by the underlying regulatory information collection requirements;

''(3) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

''(4) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

''(5) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(6) use, be consistent with, and implement applicable accounting and reporting principles.

''(c) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this section, the Commission shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.''.

SEC. 702. OPEN DATA PUBLICATION BY THE COMMODITY FUTURES TRADING COMMISSION.

The Commodity Exchange Act (7 U.S.C. 1 et seq.), as amended by section 701, is further amended by adding at the end the following:

"SEC. 25. OPEN DATA PUBLICATION.

"All public information published by the Commission under this Act shall be made available as an open Government data asset (as defined under section 3502 of title 44, United States Code), freely available for download in bulk and rendered in a human-readable format and accessible via application programming interface where appropriate.".

SEC. 703. RULEMAKING.

(a) IN GENERAL.—Not later than the end of the 2-year period beginning on the date of the enactment of this Act, the Commodity Futures Trading Commission shall issue the regulations required under the amendments made by this title.

(b) SCALING OF REGULATORY REQUIREMENTS.—In issuing the regulations required under the amendments made by this title, the Commodity Futures Trading Commission may scale data reporting requirements in order to reduce any unjustified burden on smaller regulated entities.

(c) MINIMIZING DISRUPTION.—In issuing the regulations required under the amendments made by this title, the Commodity Futures Trading Commission shall seek to minimize disruptive changes to the persons affected by such regulations.

SEC. 704. NO NEW DISCLOSURE REQUIREMENTS.

Nothing in this title or the amendments made by this title shall be construed to require the Commodity Futures Trading Commission to collect additional information under the statutes amended by this title, beyond information that was collected under such statutes before the date of the enactment of this Act.

TITLE VIII—NATIONAL CREDIT UNION ADMINISTRATION

SEC. 801. DATA STANDARDS.

Title I of the Federal Credit Union Act (12 U.S.C. 1752 et seq.) is amended by adding at the end the following:

"SEC. 132. DATA STANDARDS.

''(a) REQUIREMENT.—The Board shall, by rule, adopt data standards for all information and reports regularly filed with or submitted to the Administration under this Act.

''(b) CHARACTERISTICS.—The data standards required by subsection (a) shall, to the extent practicable—

''(1) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(2) enable high quality data through schemas, with accompanying metadata (as defined under section 3502 of title 44, United States Code) docu-

mented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

''(3) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

''(4) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

''(5) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(6) use, be consistent with, and implement applicable accounting and reporting principles.

''(c) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this section, the Board shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.''.

SEC. 802. OPEN DATA PUBLICATION BY THE NATIONAL CREDIT UNION ADMINISTRATION.

Title I of the Federal Credit Union Act (12 U.S.C. 1752 et seq.), as amended by section 801, is further amended by adding at the end the following:

"SEC. 133. OPEN DATA PUBLICATION.

"All public information published by the Administration under this title shall be made available as an open Government data asset (as defined under section 3502 of title 44, United States Code), freely available for download in bulk and rendered in a human-readable format and accessible via application programming interface where appropriate.".

SEC. 803. RULEMAKING.

(a) IN GENERAL.—Not later than the end of the 2-year period beginning on the date of the enactment of this Act, the National Credit Union Administration Board shall issue the regulations required under the amendments made by this title.

(b) SCALING OF REGULATORY REQUIREMENTS.—In issuing the regulations required under the amendments made by this title, the National Credit Union Administration Board may scale data reporting requirements in order to reduce any unjustified burden on smaller regulated entities.

(c) MINIMIZING DISRUPTION.—In issuing the regulations required under the amendments made by this title, the National Credit Union Administration Board shall seek to minimize disruptive changes to the persons affected by such regulations.

SEC. 804. NO NEW DISCLOSURE REQUIREMENTS.

Nothing in this title or the amendments made by this title shall be construed to require the National Credit Union Administration Board to collect additional information under the statutes amended by this title, beyond information that was collected under such statutes before the date of the enactment of this Act.

TITLE IX—FEDERAL HOUSING FINANCE AGENCY

SEC. 901. DATA STANDARDS REQUIREMENTS FOR THE FEDERAL HOUSING FINANCE AGENCY.

Part 1 of subtitle A of the Federal Housing Enterprises Financial Safety and Soundness Act of 1992 (12 U.S.C. 4501 et seq.) is amended by adding at the end the following:

"SEC. 1319H. DATA STANDARDS.

''(a) REQUIREMENT.—The Agency shall, by rule, adopt data standards for all information that is regularly filed with or submitted to the Agency under this Act.

''(b) CHARACTERISTICS.—The data standards required by subsection (a) shall, to the extent practicable—

''(1) render data fully searchable and machine-readable (as defined under section 3502 of title 44, United States Code);

''(2) enable high quality data through schemas, with accompanying metadata (as defined under sec-

tion 3502 of title 44, United States Code) documented in machine-readable taxonomy or ontology models, which clearly define the data's semantic meaning as defined by the underlying regulatory information collection requirements;

''(3) assure that a data element or data asset that exists to satisfy an underlying regulatory information collection requirement be consistently identified as such in associated machine-readable metadata;

''(4) be nonproprietary or made available under an open license (as defined under section 3502 of title 44, United States Code);

''(5) incorporate standards developed and maintained by voluntary consensus standards bodies; and

''(6) use, be consistent with, and implement applicable accounting and reporting principles.

''(c) INCORPORATION OF STANDARDS.—In adopting data standards by rule under this section, the Agency shall incorporate all applicable data standards promulgated by the Secretary of the Treasury.''.

SEC. 902. OPEN DATA PUBLICATION BY THE FEDERAL HOUSING FINANCE AGENCY.

Part 1 of subtitle A of the Federal Housing Enterprises Financial Safety and Soundness Act of 1992 (12

U.S.C. 4501 et seq.), as amended by section 901, is further amended by adding at the end the following:

''SEC. 1319I. OPEN DATA PUBLICATION.

''All public information published by the Agency under this Act shall be made available as an open Government data asset (as defined under section 3502 of title 44, United States Code), freely available for download in bulk and rendered in a human-readable format and accessible via application programming interface where appropriate''.

SEC. 903. RULEMAKING.

(a) IN GENERAL.—Not later than the end of the 2-year period beginning on the date of the enactment of this Act, the Federal Housing Finance Agency shall issue the regulations required under the amendments made by this title.

(b) MINIMIZING DISRUPTION.—In issuing the regulations required under the amendments made by this title, the Federal Housing Finance Agency shall seek to minimize disruptive changes to the persons affected by such regulations.

SEC. 904. NO NEW DISCLOSURE REQUIREMENTS.

Nothing in this title or the amendments made by this title shall be construed to require the Federal Housing Finance Agency to collect additional information under the

statutes amended by this title, beyond information that was collected under such statutes before the date of the enactment of this Act.

TITLE X—MISCELLANEOUS

SEC. 1001. RULE OF CONSTRUCTION.

Nothing in this Act or the amendments made by this Act may be construed to alter the existing legal protections of copyrighted material or other intellectual property rights of any non-Federal person.

○

How Data WILL Determine the Future of RegTech

Donnelley Financial Solutions

Introduction

In heavily-regulated industries, such as financial services, health care, and insurance, "RegTech" is all the rage. Every major consulting firm has released a white paper on the subject. Over just the past two years, Google searches for the term have increased sevenfold.

What is RegTech, and why is it a newly-ascendant buzzword? The Institute of International Finance defines RegTech as "the use of new technologies to solve regulatory and compliance requirements more effectively and efficiently." Meanwhile, PwC says: "Rising compliance costs, along with regulators' and the industry's growing interest in automation, have created an environment ripe for disruption by emerging RegTechs, the innovative technologies that are addressing regulatory challenges in the financial services world."

RegTech solutions include new ways to 1) automate regulatory reporting, 2) derive insights from regulatory information, and 3) share information on complex markets and products. This paper describes four examples: in the first category, Donnelley Financial Solutions' ActiveDisclosure solution; in the second, idaciti and Intrinio; and in the third, TruSet.

All four of our examples use standardized data fields and formats to create new efficiencies. All four could deliver even greater efficiencies if U.S. regulatory agencies achieved better standardization. One of the key messages here is that *RegTech solutions require data standardization*.

Conversely, if regulatory agencies fail to make further progress in standardizing the data fields and formats for the information they collect from the industries that they regulate, RegTech solutions cannot deliver much by way of further growth.

Donnelley Financial Solutions Director of Business Development John Truzzolino explains that the future of RegTech, including Donnelley Financial Solutions ActiveDisclosure, rests on "the migration from [regulatory] disclosure of documents to disclosure using structured data." Craig Clay, president of global capital markets for Donnelley Financial Solutions, explains: "When RegTech connects with data... it's a way of creating transparency. It's a disruptive approach to solutions that rests on a few key themes: efficiency, minimizing risk, and improving quality."

Hudson Hollister, founder and executive director of the Data Coalition, a Washington, D.C.-based trade association, notes that today's regulatory environment is characterized by unprecedented complexity. As of late 2017, the Securities and Exchange Commission (SEC) alone used 600 separate forms to collect disclosures from public companies, financial firms, funds, and exchanges. The SEC has transformed a few of these disclosures from old-fashioned documents into standardized data—and those disclosures are the ones that now support RegTech solutions.

If the SEC transformed all 600 of its forms from documents into standardized data, and if other federal, state, and local regulatory agencies did the same—then a true transformation would be underway.

Hollister maintains that when data is missing from the RegTech equation, the power of the solutions is greatly diminished. Hollister emphasizes, "RegTech solutions require good, structured data. It's that simple."

Some of the innovators profiled in this paper developed their ideas before "RegTech" became a buzzword. Emily Huang, CEO and co-founder of idaciti, says that when she founded her company in 2014, "the term 'RegTech' wasn't even in the vocabulary … We never said: 'Hey, we want to play in the RegTech space.'" Huang developed a new way to use standardized public-company financial data to bring insights to investors, an idea that fits the phrase that only later achieved popularity.

Like Truzzolino and Hollister, Huang maintains that more attention must be paid to the standardization of regulatory data. When she co-founded idaciti, "the missing piece was not the availability of the [information], but rather the usability of the [information as] data. And we wanted to show how the data could play an important part" in delivering easier, better insights to investors.

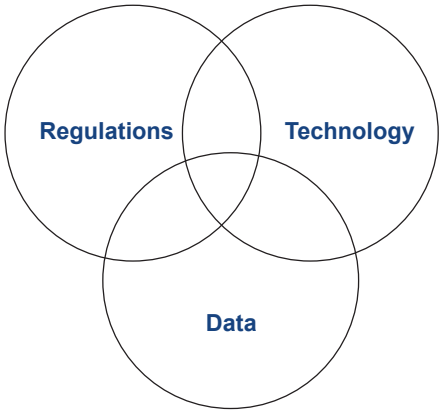
> *"When RegTech connects with data... it's a way of creating transparency. It's a disruptive approach to solutions that rests on a few key themes: efficiency, minimizing risk, and improving quality."* – Craig Clay, President, Global Capital Markets, Donnelley Financial Solutions

"RegTech solutions require good, structured data. It's that simple." – Hudson Hollister, Founder and Executive Director, Data Coalition

Key Takeaways

- RegTech solutions require data standardization.
- Regulatory agencies can maximize the promise of such solutions by coordinating changes in regulations, technology, and data.
- As structured data replaces old-fashioned, document-based disclosures, it ought to be subject to audit and quality requirements.
- In the long term, regulatory agencies can create an entirely new paradigm for RegTech by embracing a solution like Standard Business Reporting (SBR), which would serve as a common data structure for multiple regulators' regimes.

The Intersection of Regulations, Technology, and Data



If a Venn diagram were drawn with "regulations," "technology," and "data" as the three circles, the center is where new solutions scale fastest and deliver the greatest benefit. To deliver the most meaningful RegTech transformation, changes are needed in all three areas, and these changes must be coordinated with one another.

Standard Business Reporting, as practiced in Australia, the Netherlands, and elsewhere, shows what happens when government and industry work together to change regulations, technology, and data in a coordinated fashion.

Starting in 2008, the Australian government began adopting a single, standardized data structure for the information that its regulatory agencies collect from industry. In 2010, the government published its first version of a comprehensive taxonomy, or list of data fields, covering Australian companies' reports to multiple regulatory agencies. At the same time, the government worked with the Australian software industry to encourage tech companies to build software using these data fields to automate regulatory compliance. In addition, regulations were adjusted to align reporting procedures with a data-centric, rather than document-based, compliance model.

For Australia, these coordinated changes in regulations, technology, and data have created a true RegTech transformation. Australian software vendors used the standardized data structure to build new compliance solutions. Using these solutions, Australian companies can now comply with at least five different regulatory reporting regimes within one software environment. By the 2014-15 fiscal year, Standard Business Reporting was saving Australian companies over $1 billion per year through automation.

Australia's Standard Business Reporting program was patterned on an earlier, equally-ambitious program, also called Standard Business Reporting, in the Netherlands. The Dutch program also created a standardized data structure, coordinated with changes to regulations and the development of new technologies. In the Netherlands, the Standard Business Reporting regime is now the sole means by which many regulatory reports are submitted from industry to government. Meanwhile, the Estonian government has gone even further, adopting a common data structure for all government interactions by both companies and individual citizens.

The Australian and Dutch SBR programs both used the eXtensible Business Reporting Language (XBRL), a freely-available standard for exchanging business information. XBRL allows the expression of semantic meaning commonly required in business reporting. Each SBR program combined a common taxonomy of definitions used in regulatory reporting with XBRL to create a predictable electronic structure for compliance reports submitted by companies to government agencies.

For Australia, the Netherlands, and Estonia, the RegTech transformation offers more than just automatic compliance for businesses. It also dramatically improves the accuracy of information being reported because there are fewer opportunities for mistakes to be made during manual transcription. For example, once Dutch tax software vendors began using the Dutch taxonomy to automatically validate tax reports before submitting them to the Dutch Tax Administration, filers made more corrections before submission and the quality of final submissions to the agency improved.

Like its Australian and Dutch counterparts, and regulatory agencies in nearly thirty other countries around the world, the U.S. Securities and Exchange Commission uses XBRL to collect financial information. One meaningful, and long-anticipated, improvement for the SEC's regime is the adoption of the Inline XBRL, or iXBRL, format, for the disclosure forms that already involve some structured data. Inline XBRL is both human-readable and machine-readable, which means that a single document can be displayed on a browser for manual review and also ingested into software for electronic analysis. The SEC proposed a new rule to replace XBRL-based reporting with Inline XBRL in several of its disclosure forms in 2017. Final action is expected this year.

Even though Inline XBRL is much less ambitious than the multi-regulator SBR programs in other countries, it would be a good example of a well-coordinated change to regulations, technology, and data if it were formally adopted for some of the SEC's disclosure forms.

Until U.S. regulatory agencies get more ambitious, they will leave many potential RegTech solutions unexplored. Rachel Carpenter, co-founder and CEO of Intrinio, built her business with an intent to democratize financial market data access by cleaning up publicly-available regulatory data sets and making them available as easily-connected data feeds. But she says that Intrinio's ability to do this depends on "the regulatory environment increasing the amount of data sets that are filed digitally." Carpenter points out that while the SEC has mandated the use of XBRL data within some corporate disclosures, many other agencies have not adopted any data structure at all for the information they collect.

For example, if municipal governments and nonprofits—all of which generate government-mandated financial statements—made that data accessible in XBRL or even via an API, an Excel add-in, or Google Sheets, consumers could gain access to critical data in a much more flexible and open way, says Joey French, Carpenter's co-founder and President and COO of Intrinio.

French points out that hundreds of thousands of city financial reports are sitting in PDF files on tens of thousands of websites across the United States. "You can't get access to the data. Nobody can analyze a municipal bond. It's 2017 and we aren't filing that data digitally. It's insanity," says French.

"*It's 2017 and we aren't filing [municipal bond] data digitally.*" − Joey French, Co-Founder, President and CFO of Intrinio

XBRL, Inline XBRL, and SBR

- **What is XBRL?** XBRL—eXtensible Business Reporting Language—is a freely available and global standard for exchanging business information. XBRL, a specification developed and published by XBRL International, Inc., allows the expression of semantic meaning for terms commonly required in business reporting.
- **What is Inline XBRL?** Inline XBRL, or iXBRL, allows structured data XBRL tags to be included behind the scenes in a human-readable format that is displayed on a browser instead of being located in a separate document. The structured data, which can easily be processed by analytical tools, is closely tied to the numbers and text presented within the human-readable format.
- **What is SBR?** Standard Business Reporting, or SBR, starts with XBRL and other syntaxes and incorporates a taxonomy of definitions used in government legislation and reporting; these harmonized terms are then linked to the same exact standardized terms within business and local accounting software. The history of SBR began with the Netherland's Taxonomy Project in 2004, and in 2008, Australia and the Netherlands formed the SBR International Forum.

Even when regulations, technology, and data are changed in a coordinated way, data quality remains crucial.

When it comes to financial statement data, global financial markets trust the information provided because such statements are audited. According to the CFA Institute, investors are surprised when they learn that financial statements delivered digitally are not audited. Indeed, 50 percent of CFA Institute members surveyed in 2016 believe that digital information should be incorporated into the standard financial statement audit. Accordingly, Truzzolino asserts that a global move to audit iXBRL is necessary in order to ensure the trustworthiness of digital financial data. He notes that accurate, audited digital financial statements would facilitate analysis and could minimize errors in the translation from HTML to digital versions, enhancing the usability and quality of the digital data being collected.

Until recently, assumptions about the way users consumed financial statement data seemed beyond question: Users read documents, end of story. In reality, the majority of financial statement data today is consumed digitally.

When asked why financial data should be digitized, the answer is "usefulness." In a machine-readable and –consumable format, financial data can be used for far more sophisticated analysis on a company-specific basis—or across industries or even the entire population of companies out there. Doing this analysis, however, rests on data being *reliably* digitized according to approved data definitions and data standards.

Prominent voices in the industry have also argued the case for data quality. According to idaciti's Huang, "This is something I say all the time: Just because data is available doesn't mean data is usable.'"

> "*This is something I say all the time: 'Just because data is available doesn't mean data is usable.'*" – Emily Huang, CEO and Co-Founder of idaciti, Inc.

Case Studies: Four RegTech Solutions

As the last section suggests, RegTech solutions succeed when they apply technology to process data to deliver new efficiencies, in a manner consistent with regulations. Such new efficiencies can include automated regulatory reporting, faster or better insights from regulatory information, and/or shared (and therefore cheaper) information flows for complex markets and products.

The regulatory push for data quality has come as a clarion call, but specific innovations have arisen from RegTech companies themselves, including the four profiled in the following case studies.

From Donnelley Financial Solutions' (DFS's) ActiveDisclosure, which allows filers to populate regulatory submissions automatically, to idaciti, Inc., a software solution provider that offers normalized structured financial data and auto-tags additional non-financial data within financial disclosures, these emerging solutions are furthering a coordinated RegTech Data vision. Intrinio, creator of a RegTech Data marketplace, is helping disseminate machine-readable information to a broader audience, while TruSet is testing the limits of how blockchain can help a community of users contribute to a common set of machine-readable data for fixed-income prospectuses.

For now, each of these solutions takes advantage of regulatory data currently available, applying its own technological innovations, and complying with current regulations.

Donnelley Financial Solutions' ActiveLink application, for instance, pulls reporting data directly from Excel into an SEC disclosure document, promoting accuracy by eliminating error-prone reformatting, cutting, and pasting. "Clients have confidence that the numbers in the SEC filing came right from their financial reports," says Darren Peterson, DFS's Senior Software Product Leader.

Will Janensch, Co-Founder and COO of TruSet, is using blockchain technology to improve data quality by inviting a community of users to come together to "cleanse" the structured data around fixed-income instruments. "One of the keys to making any kind of automation work is that the inputs that go into the system need to be correct and trusted," says Janensch.

But if the SEC and other U.S. regulatory agencies further modernize their regulations and transform more document-based disclosures into data, while at the same time ensuring the quality of that data, many new possibilities will open up.

If regulatory agencies replace more of their disclosure documents with structured data, then "[t]he information reported to the

agencies can be used by the agencies immediately, without having to correct or change or question the data being collected," says DFS's Truzzolino. At the same time, he explains, users of RegTech solutions like ActiveDisclosure will shave hours and days off validating data and can instead "spend more time on the management of the company and the analysis of the data that's been collected."

> "*One of the keys to making any kind of automation work is that the inputs that go into the system need to be correct and trusted.*" – Will Janensch, Co-Founder and COO of TruSet.

1. Automated Reporting: ActiveDisclosure

What if a filing solution for corporate issuers that relies on good structured data could also help issuers generate higher quality data that could be validated in real time?

Since the first half of 2013, when ActiveDisclosure, Donnelley Financial Solutions' disclosure management SaaS application, was publicly released, issuers mandated by the SEC to create machine-readable data can meet their regulatory commitments more simply and more accurately, according to Peterson.

ActiveDisclosure is able to provide more than just compliance documents because the SEC decided to adopt XBRL as its standardized data format for financial statements. Should the SEC adopt standardized data for the remainder of its required corporate disclosures, then ActiveDisclosure would be able to automate an array of additional tasks that today require manual compliance.

Issuers prepare XBRL filings within ActiveDisclosure, which is a collaboration platform for finance and SEC reporting teams who prepare quarterly and annual reports. The solution ensures that important steps along the filing journey are successfully completed, explains Peterson.

It's also increasingly clear that automated reporting software can help resolve some of the problems within the RegTech Data paradigm. Take, for instance, the nagging problem of data quality. Peterson points out that ActiveDisclosure software resembles the concept of straight-through processing embraced by other industries because it is built on the notion that data should flow from the source system to the destination without the need to manipulate it by hand. He continues: "As soon as you [manually] touch data, whether it's in Excel or a desktop, there's the opportunity for errors to be introduced."

A push for data quality has direct implications for workplace efficiency, as ActiveDisclosure users can attest. "We've had clients comment that they spend as many as 800 person-hours per year 'ticking and tying' numbers to ensure they are accurate back to the source locations/systems," says Peterson. "That's one of the key reasons why ActiveDisclosure was designed to link to those sources rather than attempt to replace or replicate them."

Going forward, DFS seeks ways to integrate its existing solution so that it can help fulfill other compliance needs, as well. Peterson emphasizes that much of compliance revolves around Sarbanes-Oxley Act compliance, so DFS is now including AuditBoard's SOXHub technology, with all its references and checklists, into the process of creating SEC reports.

Ranging from private companies to government entities, any organization that produces large documents containing data and narrative content can benefit from the ActiveDisclosure platform, says Peterson. That's because ActiveDisclosure brings structure and visibility to large, multiple-contributor content projects.

2. Faster, Better Insights: idaciti and Intrinio

What if machine-readable data could be made available for parts of financial documents that are typically not tagged, allowing business insights to flourish and meaningful comparisons to be made between different companies and industries?

For Huang, the key is looking beyond the data that is currently being tagged and making all existing data available for a much wider array of purposes. By building on the SEC's existing foundation of requiring that structured data be supplied for financial statements, idaciti, a software tool that facilitates the accessing, analyzing, and visualizing of financial and non-financial data, has become a platform that can help users make sense of unstructured content, as well. Thus, idaciti uses technological innovation to extend the bounds of what is possible with currently-available, machine-readable data.

"When I read a 10-K, I go right to the MD&A [Management Discussion & Analysis]. That's where management discusses how they performed and answers the important 'why' for their performance," says Huang, noting that the SEC does not require the MD&A to be tagged. "So if you look at what XBRL has covered, it's a lot of 'what.' But the answer to 'how' and 'why' are in the unstructured portions of the documents."

Idaciti has created a software solution that uses machine learning to auto-tag unstructured portions of financial filings and capture the valuable insights in parts of a financial submission that were previously overlooked in the tagging process.

This solution also extends the existing XBRL taxonomy to key performance indicators (KPIs) that may matter within a particular industry but are not captured by the XBRL tagging process. Huang points out that revenue, which is part of the XBRL taxonomy, is not necessarily a KPI for a social media company, while the number of active users and the number of active users on mobile devices *are* valuable pieces of information (and yet the SEC has not required these metrics be tagged). When idaciti's software captures these KPIs in a machine-readable format, then meaningful comparisons can begin to be made among social media companies.

XBRL-tagged information can be invaluable when there are new accounting standards such as those around revenue recognition. While companies like Microsoft pioneered the adoption of new revenue recognition standards—and even discussed the process for doing so in SEC filings—most other companies did not reap the benefits of Microsoft's early adoption experience because the data was buried.

"When searching through all public company filings, it is impossible to locate a concept like who early-adopted revenue recognition," says Huang. "Yet companies also want to know: What kind of money and time did a company spend to adopt revenue recognition? And overall what was the impact of adoption?" Once sections of SEC filings on revenue recognition and other processes are tagged, the information will be readily available to all interested parties.

"For us, it's not just about the numbers provided to the regulator," says Huang. "A lot of narratives included in the filing truly can help the company learn from what other companies are doing and what are the best practices."

As an example of how idaciti might make a difference, Huang points out that an oil and gas company that historically spent 640 hours a year comparing its data with 25 other companies found that those comparisons could be made in a matter of minutes with idaciti. More importantly, though, the idaciti platform led to a shift in how this company benchmarked itself against peers beyond the original group of 25. Huang suggests this shift is critical: "This oil and gas company is now thinking about what possibly can be done, rather than just focusing on what it's able to do."

Huang is convinced that regulators need solutions like idaciti's platforms as a way of showing that tagging data in XBRL can have enormous strategic pay-offs. If the regulators could show the potential benefits from comparing a company to its peers or how companies are handling the adoption of new accounting standards, then they could spotlight the true value in digitizing financial data.

"With RegTech, unless you can demonstrate the benefit and quantify the effectiveness a company can get from the technology, you can't move forward," says Huang. "Knowing that 640 hours can be reduced to minutes—or the impossible task of finding the early adopters of an important accounting policy standard can be accomplished—is what needs to be showcased."

Intrinio, which has built a marketplace for over 200 data sets, including XBRL-based ones, also automates insights by allowing users to connect usable, quality data in groundbreaking ways. This RegTech solution exemplifies the value of high-quality, structured data by cleaning up existing data sets from the SEC and other sources and making them far more widely available to users.

Carpenter points out that roughly half of Intrinio's users are investors, eager to analyze financial statements, and half are developers seeking access to quality data in order to innovate by developing mobile and web apps and risk analytics software, among other things.

She describes Intrinio, which has a marketplace that has grown to 20,000 users within just the past two years, as "the Amazon of financial data." Carpenter continues, "It's a website that you can go to and shop around for whatever data you need. You have a user profile and a shopping page, where you can basically scroll through all the different types of data available and access them."

As an example of the type of innovation that Intrinio is facilitating, Carpenter cites academic research. At a quant modeling class at the University of Tampa, students were attempting to analyze the price-to-earnings ratios for 500 tickers. Doing this manually meant going to the Internet and copying p/e ratios into Excel spreadsheets for all 500 tickers. Once students had access to an Excel product available on Intrinio, they simply dragged down a formula, populating all of the data automatically. "Because of XBRL and because of these mandates, we were able to source and clean up the data and get it into the hands of students, saving them a lot of time," emphasizes Carpenter.

As another example, French cites an Alexa application designed to answer financial questions ranging from the address of Amazon's corporate headquarters to the p/e ratio for Apple. He points out that when such an application is fueled by data from a marketplace like Intrinio's, a user can basically ask Alexa any question about a publicly-traded company and receive an answer in seconds.

French believes that good, structured data must be widely available. "The Holy Grail for us," concludes French, "is that the more types of data are filed digitally, the better—at the federal level, at the municipal level, and across different asset classes."

As valuable as solutions like idaciti's and Intrinio's have become, their value would only be enhanced if the SEC and other regulators chose to continue the transformation of regulatory reports from documents into data.

"*An oil and gas company that typically spent 640 hours a year comparing its data with 25 other companies found that those comparisons could be made in a matter of minutes with idaciti.*" – Emily Huang, Co-Founder and CEO of idaciti

"*The Holy Grail for us is that the more types of data are filed digitally, the better – at the federal level, at the municipal level, and across different asset classes.*" – Joey French, Co-Founder, President and CFO of Intrinio

3. Shared Information Flows: TruSet

What if there were a way for data users to correct mistakes communally, allowing each participant in a market to spot a problem and then rectify it for all other users of the data, too?

TruSet, a start-up seeking to help bond investors share intelligence from prospectuses, is using blockchain technology to achieve that vision. "Our solution," says Janensch, "will allow you to have more accurate data that you can better trust and save money in the process."

Janensch explains that TruSet addresses what is essentially an outdated regulation: the SEC requires all issuers of fixed-income instruments—or bonds—to generate and file in HTML lengthy prospectuses that are then published on the EDGAR site. However, the institutions reading these prospectuses and investing in these bonds need the information to be presented in a structured data format, so that they can plug information into their software and run all the necessary analytics.

Enter "the middleman," or large information vendors, such as Bloomberg and Thomson Reuters, that convert fixed-income prospectuses into a machine-readable format. Janensch explains that the middlemen rely on "brute strength," automatically scraping PDFs and employing teams of hundreds of individuals who help interpret the prospectuses.

This process is expensive, and "there are enough errors that the customers of those data feeds don't trust the data as accurately representing what those prospectuses said," explains Janensch. Because of these shortcomings, he finds that each asset manager uses software analytics and back-office personnel to "interrogate the data feed" in order to locate and correct errors, creating a so-called "golden record" that they deem trustworthy. Janensch characterizes the result as "a very inefficient market ecosystem."

Fortunately, blockchain technology has the potential to up-end the paradigm. Using blockchain, TruSet has built a prototype for a solution in which a community of users corrects data that is then shared among participants.

Janensch explains that a blockchain has three important characteristics: 1) it's a distributed ledger or database; 2) there is consensus around changes made to the database; and 3) there is cryptographic immutability.

Here is how each of these blockchain features drives the TruSet solution. Although not necessarily a true distributive ledger application in the tradition of Bitcoin, TruSet's solution is "a database that's distributed among all of the different consumers of the data," says Janensch. Specifically, the community of users delves into the database, identifying errors in any reference data and only accepting the final database information that is deemed correct. Instead of participants each having to cleanse the same exact data, they work together to create a final product that they trust is accurate.

Second, consensus for TruSet takes the form of a protocol that governs under what circumstances a change to the database can occur—and whether the community can reach agreement about the accuracy of that change. "We're using that consensus process to basically crowd-source what's now currently happening in silos, which is the data correction piece happening at the customer end," says Janensch.

The third important feature is cryptographic immutability. Although the distributed database can be changed, whatever happened is still recorded within the chain. In other words, within the blockchain exists a snapshot of all of the information in the database at any given instant.

While TruSet has built its solution around the Ethereum blockhain, it is working with a permissioned (or private) version that only allows participation by pre-cleared entities. These entities become nodes, and in this peer-to-peer community, a node may play one of three roles: publisher, validator, or consumer.

One of the ironies of the fixed-income world today is that the SEC requires financial institutions to produce prospectuses as unstructured documents; however, the agency also needs machine-readable data, so it buys that data from the current large data vendors. Were the SEC to collect fixed-income prospectuses as machine-readable data in the first place, this problem would disappear.

Prototypes like TruSet are filling a gaping hole by using technology to do what the regulators are not currently doing themselves. "One of the cool things about blockchain is it allows regulators to view and maybe act on things in real time instead of waiting until after the fact to report," says Janensch.

While TruSet is debuting a fixed-income solution, the same type of blockchain solution could be created for other instruments and industries that need shared reference data and are plagued by consistency and accuracy problems.

Janensch sees enormous potential for the TruSet model because data created by financial institutions on the TruSet platform would no longer be owned by large information vendors. He anticipates a time in which a bank issuing a bond might benefit from the data it generates by getting paid a fee for contributing records. He also foresees that those validators creating golden records might no longer *pay* to have the data cleansed, but instead might *be paid* for contributing value to the overall database.

Because data vendors "own" and charge for the data they render machine-readable, vendors both charge their customers for each business function for which they use the data and impose strict usage limits on that data. Janensch hopes that this, too, might change.

The current model, he says, "has retarded innovation around the data by not allowing the community that generated the data to create more sophisticated data services." In the TruSet vision, Janensch sees fixed-income data becoming "a community-owned resource" with far fewer restrictions on how the data is used and what asset managers can do with the data that they are actively working to improve.

> "*One of the cool things about blockchain is it allows regulators to view and maybe act on things in real time instead of waiting until after the fact to report,*" says Will Janensch, Co-Founder and COO of TruSet

WHAT THE FUTURE HOLDS

In the United States, RegTech solutions will expand to the extent that the SEC and other regulators carefully coordinate changes in regulations, technology, and data. And as data replaces documents, it must be fully audited and of reliable quality.

1. The Need for Proper Auditing

In October 2017, the SEC approved a new PCAOB rule that requires significant enhancements to public company audit reports, including the disclosure of auditor tenure. These enhanced requirements became effective for audits of financial statements for fiscal years ending on or after December 15, 2017.

Arguably, the most significant change to the auditor's report is the communication of critical audit matters (CAMs), which will be mandated beginning on June 30, 2019. CAMs are matters that have been communicated to the audit committee, are related to accounts or disclosures that are material to the financial statements, and involve especially challenging, subjective, or complex auditor judgment.

"The changes adopted today breathe life into the audit report and give investors the information they've been asking for from auditors," said PCAOB Chairman James R. Doty.

Although the push for greater auditor accountability is welcome, regulators have not extended the new requirements to data submitted digitally. DFS's Truzzolino believes that overlooking structured data in this regard is a serious oversight.

Truzzolino is convinced that iXBRL could play a central role in the depth and relevance of information that users of structured data can access. A tag in iXBRL can, for instance, include a wealth of information about the data disclosed, and there is no reason why a tag should not also contain an auditor imprimatur, a link to audit guidance, and references to reported CAMs.

2. An Emphasis on Quality

Today, the financial data that companies and agencies routinely access to close their books is collected in a matter of days, and yet these companies and agencies take weeks to publish reports, delaying management and stakeholder analyses and decisions.

One reason for such delays is that information is contained in data warehouses or consolidation applications, where data is commonly cut and pasted, re-keyed, or manually transferred into word processing and spreadsheet applications.

Data standardization and the effective implementation of disclosure management applications can enhance and streamline this entire reporting process. Truzzolino notes that disclosure management applications provide report-writer functionality through word processing and spreadsheet applications commonly used in manual reporting steps. When data is standardized in this way, applications are able to pull information from disparate data sources to create automated reports.

As SEC Commissioner Kara Stein has stated, "improving the quality of data available on smaller and medium size companies could lead to improved secondary market liquidity. Improved data and transparency on market quality statistics could empower small and large investors and benefit the market overall. In short, the digital revolution is requiring us to rethink and re-envision disclosure."

In the end, RegTech Data—with an emphasis on "data"—has the potential to combine regulation/policy making, disruptive technology and data standards to streamline financial reporting, while enhancing data quality and making this digital financial data more usable for all stakeholders.

3. A Roadmap for Regulations, Technology, and Data

Following are steps that remain to be taken to achieve the RegTech Data promise in three critical realms:

Regulations. Truzzolino notes that custom extensions to the SEC's US GAAP taxonomy created by individual companies make it difficult for meaningful data comparisons between the various companies out there. He says that until there is true standardization within the structured reporting, achieving the original SEC vision "of leveling the playing field between companies large and small" for how information is presented and consumed by investment analysts will remain an elusive goal.

What's more, the SEC needs to fix the lack of comparability across its current XBRL-formatted financial statement submissions, contends Hollister. He is eager for "the SEC and

other regulators [to] adopt a standardized data structure for all of the information they collect from the financial industry."

Technology. Intrinio's Carpenter and French point out that software developers attempting to devise new solutions need quality data in a machine-readable format in order to build the solutions of the future. Without this data, innovation will be stifled.

Huang maintains that when new applications are introduced, it is critical that the practical uses be highlighted. As a cautionary tale, she notes that when the SEC unveiled its inline viewer for iXBRL, the regulator showcased the metadata, what she calls "the really geeky stuff," rather than the functionalities that could truly help investors.

The inline XBRL viewer enhanced by idaciti allows a user to click on a revenue number for a given company and then see not only the revenue reporting for that year—but a trending chart depicting whether revenue is increasing or decreasing. Huang notes that the inline viewer also makes it easy to benchmark any reported item for several companies at once. "The inline viewer brings the data to life, and people can see that this is the power of XBRL," she says. "The SEC has created a great foundation, but it's up to the agile software companies in the marketplace to add additional, innovative functionalities."

Huang is adamant that all technology should be presented in a way that makes the value to users clear. "It's important to show how we maximize the value and effectiveness for the issuer by automating using machine learning and other advanced technologies," she says. "We want to show how you can ask a question and get an answer without taking 17 steps to get there."

Data. As Hollister points out, many RegTech Data applications "are held back by a lack of accurate data." Making machine-readable data dependable is essential for true progress to be made.

When Intrinio's Carpenter envisions the RegTech Data future, she says that the goal should be "a superhighway" in which reliable and accurate data is generated by companies and flows instantaneously to users and innovators alike. With this data, companies, investors, and software developers will have the infrastructure necessary to gain needed insights for breakthrough developments.

In the short term, this means the SEC should adopt the Inline XBRL format for the financial statements that it already collects in XBRL. While in March 2017 the SEC signaled that it will mandate iXBRL as the reporting language for financial statements, that requirement has not yet taken effect. The SEC continues to require an HTML copy (in a human-readable format) and a separate exhibit in XBRL (as the machine-readable format).

Once an iXBRL mandate occurs for public companies in the United States, then those companies will no longer have to submit to the SEC the HTML and XBRL versions of the exact same filing. This shift will take some of the burden off issuers themselves, and it will result in greater accuracy because there will be a single filing rather than two formats that could have discrepancies between them.

In the medium term, the SEC and other regulators need to replace existing disclosures with standardized and structured data, encourage the development of both public-sector and private-sector technology to take advantage of that data, and adjust regulations to permit all manner of compliance to take place in a data-centric, rather than document-based, manner.

In the long term, many industry experts are eager for regulators to adopt an even broader and more far-reaching standardized data structure, such as Standard Business Reporting. Hollister points out that SBR would allow data collected from all corners of the government to be used in meaningful comparisons, spurring dramatic new RegTech applications.

> "*We want to show how you can ask a question and get an answer without 17 steps in the middle.*" – Emily Huang, Co-Founder and CEO of idaciti.

4. A Future Defined by Data

Hollister and Truzzolino point out that with the Financial Transparency Act and other Congressional developments afoot, now is a propitious moment to take a closer look at how data collection can be standardized across government agencies.

Meanwhile, Hollister notes that the Data Coalition is seeking reforms in Congress that would require financial regulators to adopt standardized data formats across all the information that they collect. Specifically, he hopes to see SBR adopted in the United States as it has been in the Netherlands and Australia.

"SBR," says Hollister, "is a standardized data structure that the United States should adopt for all the information that companies report to regulatory agencies." He continues: "Writ large, SBR is the culmination of RegTech Data. It's not possible to scale RegTech Data solutions unless you have data standards, and that's why we are pursuing the adoption of data standards by government regulatory agencies."

The goals of RegTech Data and SBR are sweeping but increasingly within reach. And while SBR looks promising, it is not the specific solution that matters so much as the overall commitment by regulators to demanding structured data that is presented in a consistent and easy-to-use way. The move from documents to structured data appears inevitable, but regulators will continue to play a critical role in helping achieve the full potential of the RegTech Data vision.

In the coming months and years, what will almost certainly be central to the success of RegTech Data is ensuring that "data" becomes an integral part of all facets of the RegTech conversation.

"Talking about 'Reg' and 'Tech' isn't enough," concludes Hollister. "Data is key."

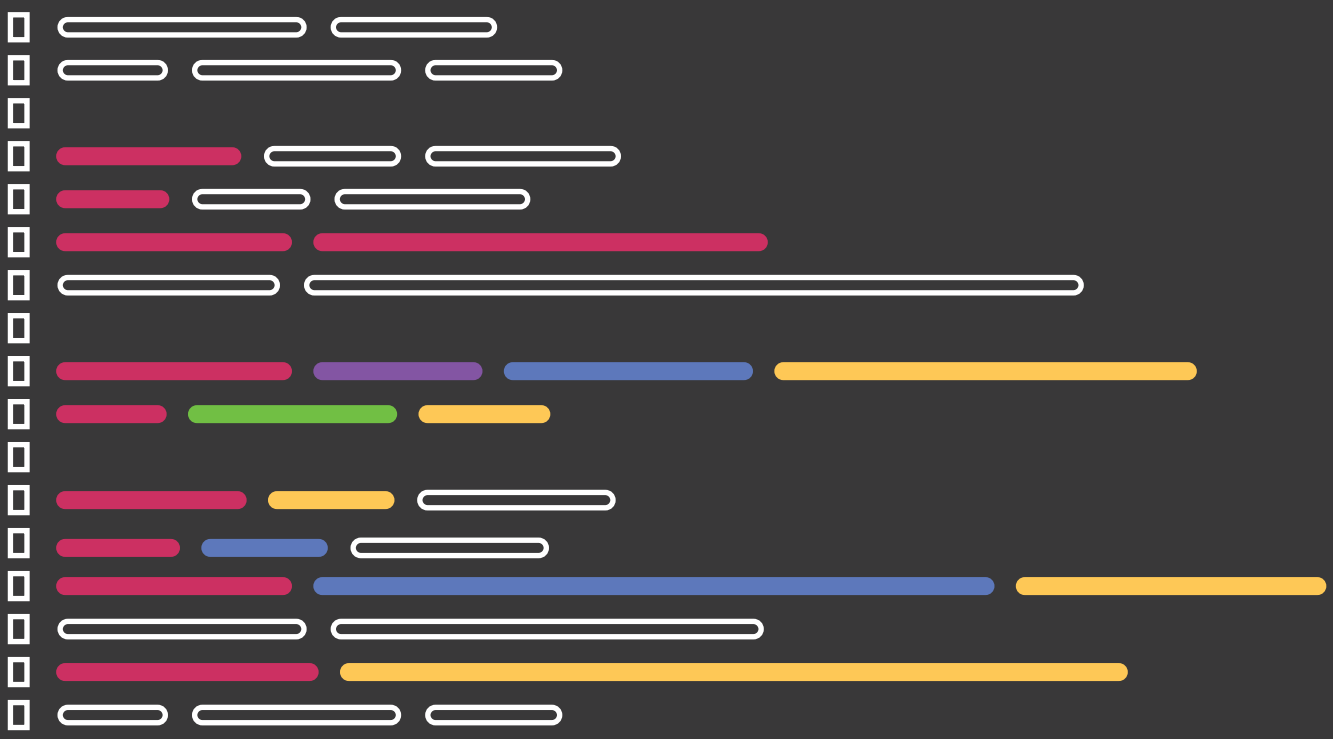
WP-RegTechData 02.2018

Corporate Headquarters
35 West Wacker Drive
Chicago, IL 60601

800.424.9001

Donnelley Financial Solutions

Understanding Machine-Readability in Modern Data Policy



DATA FOUNDATION

Workiva

Author

Dean Ritz
Board Member, Data Foundation
Senior Director of Digital Reporting Strategy, Workiva

Acknowledgements

The author wishes to thank Christina Scheuer for substantial drafting assistance; to Workiva for funding this paper; and to the following people who reviewed and commented on later drafts: Rafael Gonçalves, C. Christina Ho, Joseph Howell, Mike Starr, Richard Szabo, Catherine Tsai, and Carol Volpe.

Disclaimer

This paper is a product of the Data Foundation and Workiva. The findings and conclusions expressed by the author do not necessarily reflect the views or opinions of the Data Foundation, its funders and sponsors, or its board of directors.

Executive Summary

Application of data standards in our society offers potential to ensure potential data users have access to information about what data means. With clear delineation of data meaning, researchers, analysts, and other data users can effectively use data to create value for the economy, policymakers, and society. Data standards promote common understanding and agreement on access to information. When made machine-readable—a format that can be accessed and used by computers—the efficiency and effectiveness of data analysis approaches are enhanced. The application of data standards and machine-readable data-reporting processes, such as with regulatory or compliance reporting, is currently underway as disclosure modernization sweeps through industry, academic, and governmental entities.

This paper provides an overview of why machine-readability matters and asserts that public policy for disclosure modernization should specifically require that both the values that entities report and the compliance standards for reporting those values be structured as machine-readable data. It concludes by offer three policy recommendations (detailed on pages 23 and 24):

RECOMMENDATION #1:

Policymakers should require machine-readability when possible.

RECOMMENDATION #2:

Policymakers should clearly communicate intent in legislative and regulatory actions on the role, purpose, scope of detail rendered as data, and applicability of data standards.

RECOMMENDATION #3:

Policymakers should encourage the adoption and use of open, consensus standards to encourage cooperation, efficiency, and innovation when drafting new data policies.

Disclosure modernization supports the proper functioning of compliance and financial systems, and most importantly, of governments themselves. The benefits and practical implications are vast, including that standardization builds confidence in compliance and financial systems which quickly and reliably detect fraud, errors, and other concerns. Improving public policies related to disclosure modernization can support efforts to enhance transparency and accountability in our society, ultimately including improving public trust in institutions.

Introduction

"1" is a symbol for the numeral one. Why?

It is because we have generally agreed to it, and that agreement generates value. Analogously, data has its greatest utility when there is agreement as to what it represents. Only then can we access what data means and make effective use of it. Data standards formalize these agreements. With data standards, we can make the information that is important to us (e.g., a trial balance, financial statements, scientific test results, inventories) accessible in a machine-readable form, that is, in a format that can be easily processed by a computer and, therefore, potential users. This paper first will detail how the quality of data standards determines the effectiveness of data; second, it will explain how to elevate machine-readability requirements to maximize data value; and third, it will offer recommendations to policymakers, particularly those engaged in the areas of legal and regulatory compliance, and evidence-based policymaking.

Disclosure modernization is the name given to the transition from silos of paper and e-documents (such as the popular PDF analog of a paper document) to machine-readable data. It is a movement with active participation by policymakers worldwide. The **first wave** of this movement captures as machine-readable data the values contained in compliance reports. The **second wave** captures as machine-readable data the compliance standards[1] as well—i.e., the description of compliance requirements and authoritative definitions of the standards as machine-readable data. Without both, governments have only half of the solution, and are modernizing only part of the available knowledge.

This paper asserts that public policy for disclosure modernization should require that both the values that entities report and the compliance standards for reporting those values be structured as machine-readable data.

Machine-Readability in Theory and Practice

Background Criteria for Effective Machine-Readable Data

The key to transmitting information successfully is to package it in such a way so as to ensure the end user accepts, interprets, and uses it as intended. Information theory, which provides the conceptual foundation for machine-readable data, reframes information transmission as three levels of communication problems: (1) the technical problem of how information is transmitted from a sender to a receiver; (2) the semantic problem of how precisely a transmission conveys intended meaning; and (3) the effectiveness problem of the transmission bringing about the desired end.[2]

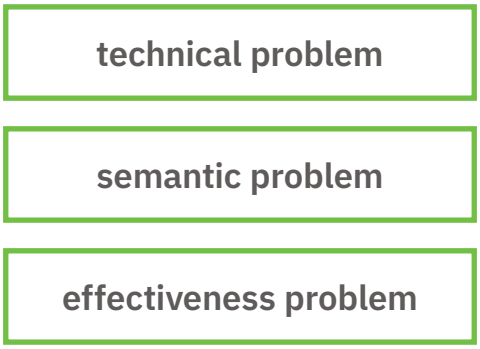


Illustration 1: The 3 levels of communication problems.

The effectiveness of communication is dependent on how the technical and semantic issues are addressed. Fortunately, the technical problems associated with transmitting financial data have been solved and already are applied in practice. We have expressive data encoding languages (e.g., XBRL®, XML, JSON, RDF, etc.[3]) complemented by scalable, secure, high-performance data transmission and storage systems. This leaves us with the semantic problem for machine-readable compliance reports, and accordingly, the possible semantic solutions are the focus of this paper. The criteria for evaluating semantic solutions are scope, precision and completeness.

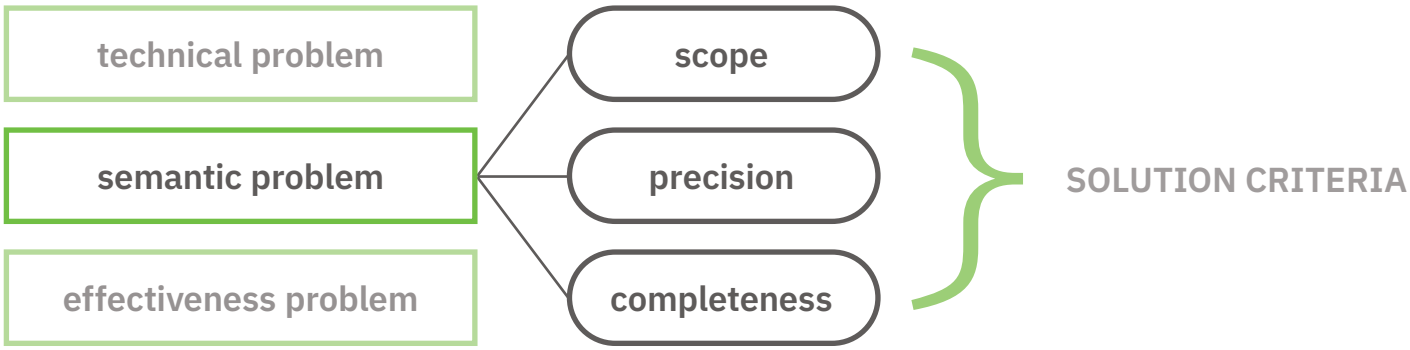


Illustration 2: Criteria of a semantic solution

First, let's take the first wave of disclosure modernization where the content of compliance reports are delivered as machine-readable data. In this context, **scope** is the universe of what is required to be reported. **Precision** is the unambiguous meaning of what is reported. **Completeness** is reporting of all the data that is applicable.

The second wave requires that compliance standards also be represented as machine-readable data, enabling computers to validate that a report meets requirements for scope, precision, and completeness. This knowledge-driven automation by machine makes the transition—from e-document and paper reports to machine-readable data—easier for the preparers of those reports, and easier for the consumers of those reports, too.

> **It is the second wave of disclosure modernization that most reduces the cost and burden of compliance reporting.**

To capture both the first and second waves of disclosure modernization we must broaden the definition of the criteria beyond the first wave "what is required to be reported." Scope now pertains to the breadth of details covered by the information reported and the compliance standards for the report.

In both the first and second waves of disclosure modernization precision remains focused on exactness. Precision eliminates ambiguity and creates consistency. However, precision is not the same thing as accuracy, as one may lie with precision. Ironically, there can be value in working with inaccurate but precisely described data (e.g., anomalous data may indicate fraud or a faulty measuring device).

Lastly, completeness is capturing all of the information required to be effective. In our focus on the second wave of disclosure modernization, this information consists of compliance values and the compliance standards that together bring about automated validation of data quality.

Here is an example of how this works. All accountants know the standard accounting equation of Assets = Liability + Equity. This is knowledge about the domain of accounting. The following example reports values for that equation for two time periods.

YEAR	ASSETS	LIABILITIES	EQUITY
2018	$12,000	$4,000	$8,000
2019	$13,500	$7,500	$4,000

Illustration 3: Simple example of scope, precision, and completeness

From the point of view of a 'first wave' person reading this small report, its contents have scope: a value is provided for Assets, Liabilities, and Equity. It also has precision: each has a numeric value, a measurement date, and a unit of measure. And, in the 'first wave' person's interpretation, who we assume knows the accounting equation, it is complete because this person can apply that domain knowledge to the report and see that the equation isn't satisfied for the period of 2019.[4]
A machine-readable version of these values (i.e., a first wave disclosure modernization report) would not be able to detect the error because the domain knowledge isn't also machine-readable.

To be explicit, here is a simplistic machine-readable version of the report in Illustration 3.

<Assets> <Value>12000</Value> <Unit>USD</Unit> <Year>2018</Year> </Assets>	<Liabilities> <Value>4000</Value> <Unit>USD$</Unit> <Year>2018</Year> </Liabilities>	<Equity> <Value>8000</Value> <Unit>USD</Unit> <Year>2019</Year> </Equity>
<Assets> <Value>13500</Value> <Unit>USD</Unit> <Year>2019</Year> </Assets>	<Liabilities> <Value>7500</Value> <Unit>USD</Unit> <Year>2019</Year> </Liabilities>	<Equity> <Value>4000</Value> <Unit>USD</Unit> <Year>2019</Year> </Equity>

Illustration 4: Simple example as simplified machine-readable data

What's missing from the table in Illustration 4 is the accounting equation, which in our simplified machine-readable language could look like this:

```
<ValidationRule>
        <TotalElement>Assets</TotalElement>
        <ContributorElement>Liabilities</ContributorElement>
        <ContributorElement>Equity</ContributorElement>
</ValidationRule>
```

Illustration 5: Validation rule as simplified machine-readable data

Without the accounting equation, the error is not likely to be detected by a computer. With it, the error will be detectable by a computer.

We now arrive at a bigger picture of the requirements for a semantic solution. With regard to completeness, the data should cover both assertions of fact (i.e., the numbers and their classifications as seen in Illustration 5) and also the assertions of domain knowledge (i.e., compliance rules). These two kinds of assertions together comprise what data scientists refer to as a knowledge-base.

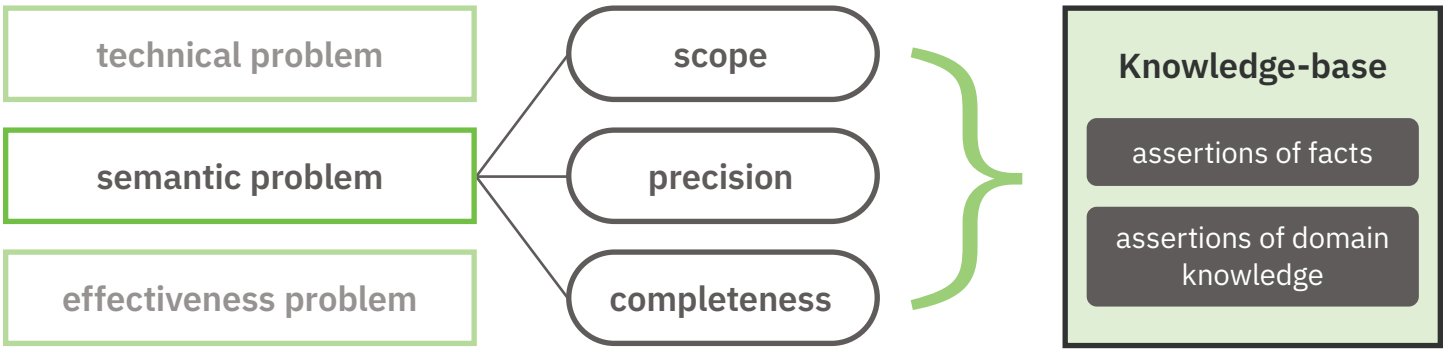


Illustration 6: Composition of a knowledge-base

Finally, regarding computer solutions to semantic problems: Computers have radically improved at appearing to 'know' things. Anyone who has voiced a question to a device (e.g., Amazon Alexa or Google Home, or nearly every smartphone) and received a correct answer has direct experience of this. And while generalized machine intelligence is still some time off, artificial intelligence opportunities for policymakers are substantially easier to achieve because there are specific knowledge domains for compliance reports, and these often are well structured.[5] The domain knowledge of interest is codified by domain-specific institutions e.g., the SEC, FERC, GASB, FASB, MSRB, EPA, FDIC, etc.[6] Domains have bounded knowledge, and there are human standards for what to report and when to report. Therefore, it is possible to achieve the levels of agreement as machine-readable data standards that will solve the semantic problem and make compliance reports much more effective. To illustrate this assertion, "An Illustrative Example" presents, and Appendix A dissects in detail, an example financial report to illustrate the kinds of domain knowledge that one would want to present as machine-readable data and, by doing so, the standards one must satisfy for machine-readable levels of agreement.

An Illustrative Example

For the benefit of public policy making and performance, and as a general recommendation, the federal government should learn from the data practices of life sciences and intelligence services—domains focused on gathering and analyzing knowledge such as genomic information and intelligence reports. These fields' charter to protect and save lives increases the necessity for solving communication problems.

> **What we should singularly learn from the data innovators in life sciences and intelligence services is that data effectiveness increasingly depends upon sophisticated representations of information**

Clinical trials and intelligence activities produce enormous quantities of data; the effectiveness of the workers in both of these fields relies on the scope, precision, and completeness of the data, ideally governed through automated processes. What we should singularly learn from the data innovators in life sciences and intelligence services is that data effectiveness increasingly depends upon sophisticated representations of information i.e., through more expressive levels of agreement that address the following premises:

- **Data producers cannot anticipate all possible uses of their data.**
- **Data users may possess additional knowledge about someone else's data.[7]**
- **In a given domain, human-readable standards are the foundation for the data standards.**
- **Data automation is the primary means to managing the increasing volume, velocity, and complexity of data.**
- **Evidence-based policymaking is expected by citizens, and is in fact the law of the land.[8]**
- **Like all communication, the evidence used to evaluate activities and work product is effective only if the technical and semantic problems are solved.**

What follows is a financial reporting example that will show how this works in practice. The example comes from the Comprehensive Annual Financial Report (CAFR). Nearly 40,000 CAFRs are produced each year by state and local governments and special districts. It is likely more familiar to readers than data from genomic studies or intelligence gathering. The CAFR plays a critical transparency role in the $3.8 trillion municipal bond market,[9] one reason to use the CAFR as an example. Secondly, nearly 30,000 of the 40,000 are provided to the federal government as part of single audit procedures for entities receiving federal grants or awards. Third, moving to machine-readable data for CAFR could reduce by $30 billion the cost of borrowing by state governments, local governments, and special districts.[10]

This particular example is from the 2019 CAFR from the State of Georgia.[11] It is considered an excellent example of a CAFR because of the breadth and clarity of this complex 401 page report. This annotated illustration, further detailed in Appendix A, makes observations as seen through the eyes of a subject matter expert (SME) reading the Net Position summary of the Statement of Net Position, both in terms of observations of fact and also the assertions of domain knowledge, i.e., the knowledge that experts bring to the example. Collectively, this is the knowledge-base we want to capture in a machine-readable form. Today, this knowledge-base is inferred by examining a visual presentation of the report.

	A	B	C	D	E
1	**STATE OF GEORGIA** 1				
2	Statement of Net Position				
3	June 30, 2019 2	**Primary Government** 5			
4	(dollars in thousands) 3	**Governmental Activities**	**Business-type Activities**	**Total** 6	**Component Units**
5	**Net Position** 4				
6	Net Investment in Capital Assets (1)	20,361,680	8,429,136	25,566,212	3,534,685
7	**Restricted for:**				
8	Bond Covenants/Debt Service	64,016	–	64,016	94,454
9	Capital Projects	–	13,076	13,076	217,230
10	Guaranteed Revenue Debt Common Reserve Fund	53,766	–	53,766	
11	Loan and Grant Programs	–	–	–	1,844,780
12	Lottery for Education	1,354,630	–	1,354,630	–
13	Motor Fuel Tax Funds 8	3,508,961	–	3,508,961	–
14	**Nonexpendable:**				
15	Permanent Trust	–	181,016	181,016	2,398,751
16	Other Programs	–	–	–	46,977
17	Other Benefits	–	305,877	305,877	–
18	Other Purposes	1,293,746	313,732	1,607,478	407,868
19	Permanent Trust Expendable	–	–	–	804,556
20	Unemployment Compensation Benefits	–	2,535,856	2,535,856	–
21	**Unrestricted (1)**	(7,660,565)	(6,201,340)	(10,637,301)	2,295,805
22	7	**$ 18,976,234**	**$ 5,577,353**	**$ 24,553,587**	**$ 11,645,106**
23	**(1) Refer to Note 4 for additional details**				

Illustration 7

We start with the first part of assertions of fact: the values.

1 **Each of the 59 numeric values is reported for the government entity called "State of Georgia."**

2 **Each is measured at the end of the fiscal period, June 30, 2019.**

3 **Each is reported in thousands of dollars, presumably rounded to the nearest thousand.**

4 **All 59 values are classified as net positions. Net position is further classified as one of either restricted and unrestricted funds.**

5 **The values in column B are further classified as "Governmental Activities." The values in column C are further classified as "Business-type Activities." The values in column D are further classified as "Total Primary Government."**

6 **The value of column D should be the sum of the values reported for columns B and C in each respective row (i.e., rows 6, 8-13, 15-16, 17-20). That is, the value reported for Primary Government should be the sum of the value reported for Governmental Activities and Business-type Activities. Take particular note of cell D6. It should equal the sum of B6 and C6. It does not. Similarly, the value of D21 should equal the sum of B21 and C21. It does not. A subject matter expert would notice that they 'don't add up' and would have to investigate further, following the trail that continues through the footnote at the bottom left of the table.**

7 **The value of row 22 should be the sum of the values above it in each respective column (i.e., columns B, C, D, and E).**

8 **Further classification of net position are found in cells A6, A8:A13, A15:A20, and A21, for the values on each label's respective row (e.g., "Net Investment in Capital Assets," "Bond Covenants/Debt Service," "Capital Projects," etc.).**

This inventory is partial, and would be complemented by the dozens of reporting requirements not mentioned but which must be addressed by those creating or reviewing this report. There are requirements regarding completeness: does the CAFR contain a Statement of Net Position (SNP); does the SNP include all the necessary parts for that statement: assets, deferred outflows, liabilities, deferred inflows, and net position; is the entity identified, and does it comply with the identification requirements?[12] There are requirements regarding consistency: are all the values reported for the same date; and does that date coincide with the end of the fiscal reporting period for the entity?

By making a CAFR machine-readable in both the reporting of facts and compliance standards, these observations could drive automated validation of correctness and completeness. Project the few observations from one page of a 401 page report, and multiply by the 40,000 produced annually, and one may get a sense of the burden carried by manual efforts to review and audit.

As of this paper's publication, only one state or local U.S. government has published on its official government website a machine-readable CAFR (or portion thereof). In 2019, Will County, IL, became the first to do so.[13]

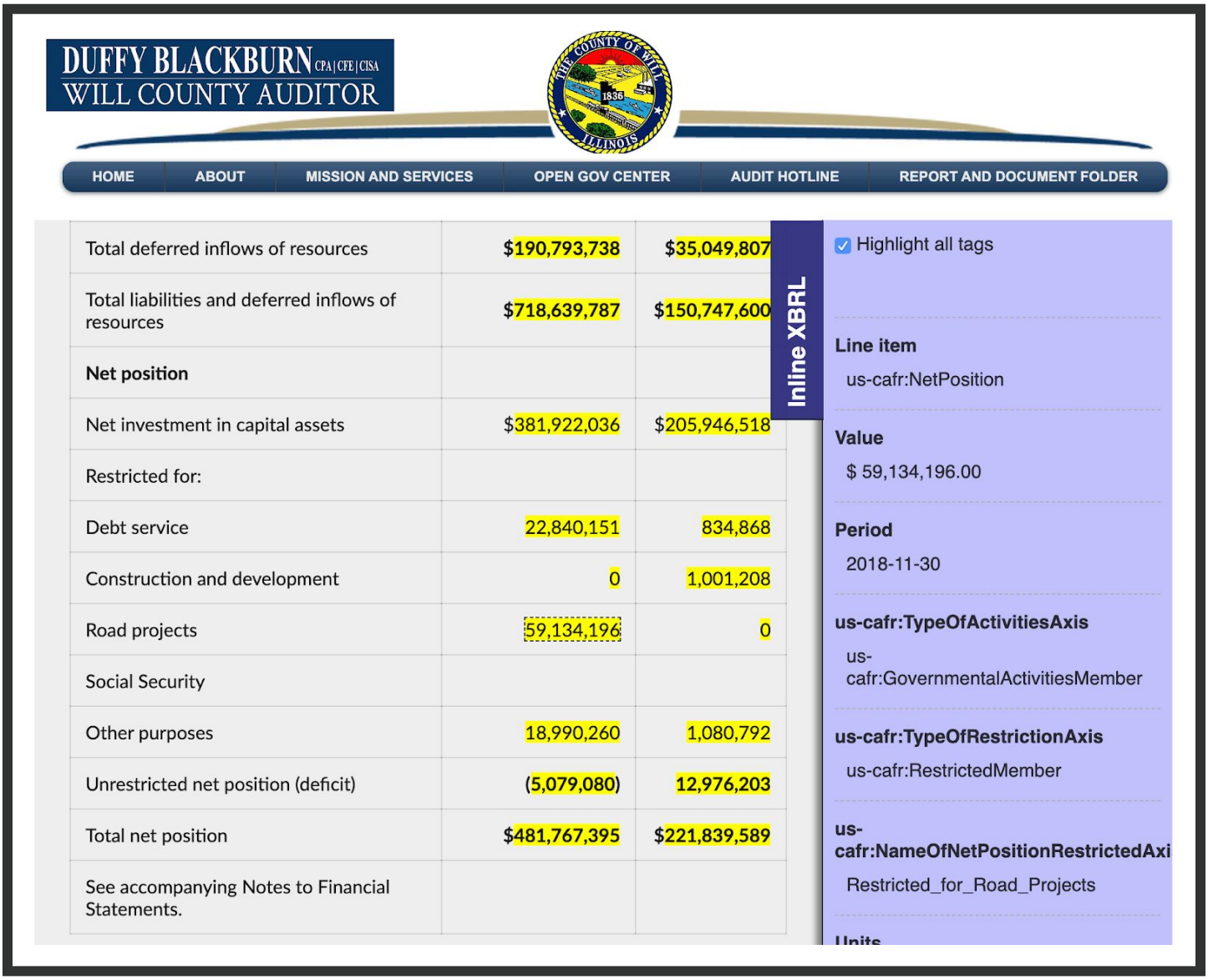


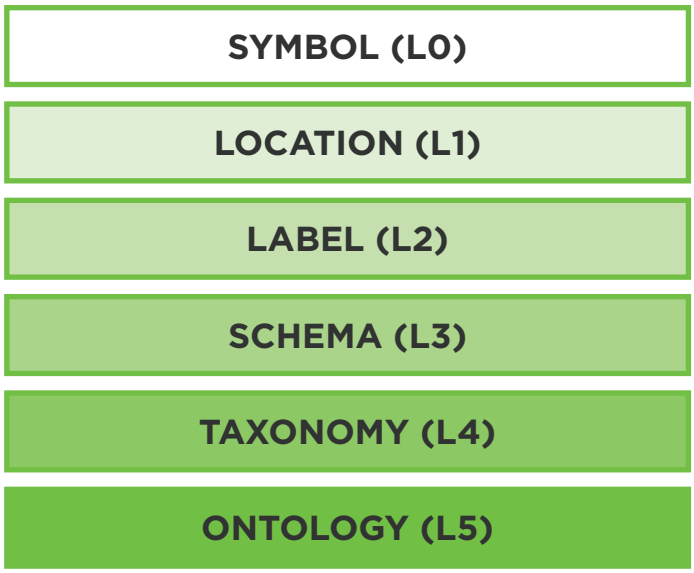
Illustration 8: Portion of the Will County, IL CAFR in an iXBRL open data viewer

While other states and local governments have considered doing this, public policy does not require it.

The Levels of Agreement for Machine-Readability

Below, we introduce names and brief descriptions of the levels of agreement that may represent the assertions of fact and the assertions of domain knowledge expected from machine-readable data. These are the six levels of agreement, arranged in order from least to most expressive:

Each is described below.



LEVEL 0 (L0): AGREEMENT AS SYMBOL.

This level of agreement is fundamental. It is an agreement that a discrete symbol exists as a unit of information. This unit of information exists prior to interpreting it as meaning something. It is simply a symbol that embodies potential communication that we shall call a datapoint.[14] Let's use the symbol "1" as an example. The person that interprets the "1" symbol as a number is making a small semantic leap based on cultural agreement; its number-ness is not intrinsic to the symbol. If the symbol is "one," a larger semantic leap is required to interpret it as a number. The effectiveness of this level of agreement is wholly dependent on the receiver/interpreter of the communication. The symbols that occupy every cell of Illustration 7 express the L0 level of agreement. Even the lines are symbols, and included within L0; the single bar above and the double bars below row 22 in Illustration 7 has meaning to accountants i.e., together they indicate a total.

LEVEL 1 (L1): AGREEMENT AS LOCATION REFERENCE.

This level of agreement captures a means to refer to the physical[15] location of an L0 datapoint. This is common in spreadsheets where references to data are represented as the intersection of a column and row coordinate (e.g., "D13" where the alpha character(s) represent the column and the numeric character(s) represents the row). The interpretation is system-dependent. For example, a spreadsheet may interpret "=D13" as the value stored in that cell location, whereas a document editor may interpret "=D13" as a piece of text comprising four characters. Some systems permit

an L1 reference to include a file system path (e.g., "[../budget/spreadsheet2.xlsx] workbook2:D13"). An L1 reference identifies a datapoint, and uniquely so, but offers no clue as to what will be found there. Every cell of Illustration 7 may be covered by an L0 location reference e.g., the L0 datapoint "13,076" may be referenced by the L1 of "D13."

LEVEL 2 (L2): AGREEMENT AS LABEL.

With this level of agreement, we finally move from an L0 datapoint as a symbol whose meaning is wholly up to interpretation, to that of having a symbolic clue as to the meaning of the datapoint. The meaning is captured in a label. It is most commonly used as a column or row header of a spreadsheet or table, communicating a convention to human readers to read the label into values in that same row or column (e.g., Illustration 3 where the table's columns are labeled "Year," "Assets," "Liabilities," and "Equity").

L2 is in widespread use through the comma-separated values (CSV) data encoding format. CSV files are familiar to almost every spreadsheet user. L2 also is an essential part of relational database management systems (RDBMS). These database systems organize data into tables where each row represents related L0 data points and each column has an L2 label. The label is used by the RDBMS to navigate to the datapoint of interest; a user does not have to know the ordinal position of the desired column as the label alone is sufficient. It is a bonus if that label actually means something by common agreement between the producer and the consumer of the data (e.g., "First Name" and "Surname" in your contact manager).

However, the move from L1 to L2 isn't so simple. Continuing with the example from L1 where "D13" locates the datapoint "13,076", it is immediately clear that the visual arrangement of the table is not shaped for such a simple singular reference because that datapoint has multiple possible labels due to it having multiple classifications: as pertaining to "Capital Projects," a "Business-type Activity," and a "Total" (i.e., "Primary Government"). The table in Illustration 7 is arranged for human eyes rather than machine-readable data. This leads us to an important point: machine-readable data can liberate data from the gridlock of spreadsheets. L2 does not get us there, but the more expressive levels of agreement do achieve this liberation.

LEVEL 3 (L3): AGREEMENT AS SCHEMA.

A schema organizes information as a hierarchy, often with an L2 label for each entry in the outline. Below is an example in the computer syntax XML. The data points and their meaning and their relationship between the data points should be obvious even if one does not know the XML syntax.

And just as an outline nests information, a schema provides for the nesting of terms (e.g., the term "name" is part of the term "person"; and terms "first" and "surname" are parts of the term "name". With the introduction of L3 it becomes possible to see how a more expressive level of agreement leads to data consistency. For example, restrictions can be expressed to require a datapoint for "born" but optional for "died". One may begin to see how L3 is necessary to represent the knowledge-base of

a CAFR e.g., the CAFR must include a statement of net position, it must include all the necessary parts for that statement, the report must identify an entity, etc. With a machine-readable knowledge-base, software programs may enforce data compliance.

```
< person>
        <name>
                <first>Claude</first>
                <surname>Shannon</surname>
        </name>
        <lived>
                <born>1916</born>
                <died>2001</died>
        </lived>
</person>
```

LEVEL 4 (L4): AGREEMENT AS TAXONOMY.

In L3 we use schemas as a means for general purpose representation of structured data. In L4, taxonomy, we use schemas to create a specific and deliberate representation of a classification system, i.e., for each datapoint we can answer the question: "Which box does this go into?" Those who remember high school biology may recall the following taxonomy of organisms: kingdom, phylum, class, order, family, genus, species. One may also recall that the application of a taxonomy usually defaults to applying the most specific classification possible.

A taxonomic specification describes a closed world assumption of what can be known, i.e., what can be put into a domain's knowledge-base. In the domain of financial reporting, each financial reporting standard has a human-readable system for classifying the contents of financial activities in a closed world system. In the U.S., the Financial Accounting Standards Board (FASB) breaks financial activities into the broad classifications of assets, liabilities, equity, comprehensive income, investment by owners, distribution to owners, revenues, expenses, gains, or losses.[16]

The FASB also maintains an implementation of the United States Financial Reporting Taxonomy (USFRT) in the XBRL data encoding language. This taxonomy is used to construct and transmit to the SEC machine-readable versions of quarterly and annual reports by all U.S. publicly traded companies. This has amounted to more than 270,000 reports in the past decade. The International Accounting Standards Board (IASB) is among the many other financial reporting authorities that maintain a financial reporting taxonomy in a data encoding language; this provides for the exchange of machine-readable financial reports between regulated entities and oversight bodies.

A taxonomy can have multiple hierarchies i.e., multiple classification schemes, one for each kind of organizing relationship. To illustrate, a taxonomy based on our CAFR example provides for relating the classification of types of net position in a hierarchy of broader-to-narrower classification schemes. The CAFR also has a hierarchy of mathematical relationships. These connections between classifications help achieve

data quality because they can express the classifications that are permitted to go together, as well as those which are prohibited from going together. In this specific case, and assuming that the GASB guidelines permit this, a GASB taxonomy could explicitly permit data points classified as Capital Projects to be optionally classified as restricted or unrestricted. Alternatively, it instead could require explicit classification as restricted or unrestricted.

Mathematical connections may be treated similarly in a summation scheme i.e., relationships that express how classifications relate by way of summation of numeric values. For example, a taxonomy could express the requirement of assertion 6 i.e., the value reported for Primary Government should be the sum of the value reported for Governmental Activities and Business-type Activities. These additional constraints further drive classification, and can be used by software to validate the reported data as compliant with the L4 taxonomy and so, at least as modeled in the taxonomy, correct. The expressiveness of an L4 taxonomy in capturing assertions of domain knowledge is part of the reason why taxonomies and suitably expressive data encoding languages have been used by FASB and the IASB for their disclosure modernization work.

LEVEL 5 (L5): AGREEMENT AS ONTOLOGY.

Given the expressive power of a taxonomy, what could an ontology offer? It offers us two significant differences. First, it permits us to specify characteristics about a taxonomy that cannot be described in the taxonomy. Second, an ontology accommodates two key premises of the open world assumption; specifically, data producers cannot anticipate all possible uses of their data, and data users may possess additional knowledge about someone else's data. An ontology provides flexibility and extensibility to the closed world system of the taxonomy.

First, we focus on how the kinds of agreements captured in an ontology accommodate the open world assumption. An open world assumption is managed by focusing on the properties of data rather than its classifications because classifications may change when data is put to new uses, or when additional information becomes available and is added to the knowledge-base. With a taxonomy, we start with the classification and unpack its properties. In an ontology, we may start with the properties and conclude with a classification. In a taxonomy we ask, "Into which box does this thing go?" In an ontology we ask, "What's in this box?" In a taxonomy, we know a sparrow is a kind of bird. In an ontology, we would look into our 'bird box' and find a sparrow.

An ontology concerns the nature of being, specifying the properties that make a particular thing (e.g., 'animal_21') a member of a particular set or multiple sets of things (e.g., a bird). Continuing with our sparrow, an ontological view of birds would identify properties such as:

- **is an animal**
- **lays eggs**
- **breathes air**

- **has wings**
- **has beak**
- **does not have lips ("has lips" is mutually exclusive of "has beak")**
- **has exactly two feet**

This is one ontological description of 'bird-ness.' In data science, an ontology is an "explicit specification of a conceptualization."[17] With our specification of bird-ness an item will be classified as a bird when its properties come to satisfy that specification. That item would then be in our 'bird box' whether or not someone else has classified it as a bird. If we had an ontological description for flying things we suspect that we would find that all things which satisfy 'bird-ness' also satisfy 'flying-things-ness,' and this is a member of more than one set of things (in contrast to a taxonomy where something may be classified as one thing only). Since the open world assumption permits new knowledge to come into the world the classification of things may change. The adaptability of L5 to represent multiple points of view as logical expressions is why ontologies are the conceptual ground for the semantic web.

An ontology can help specify machine-readable information that can be used to validate the quality of a taxonomy. An ontology allows us to express constraints or requirements on items in our knowledge-base, including the data that describes a taxonomy. For example, one would think as problematic a taxonomy that enables a data point to be classified as two different things at the same time e.g., as an asset and a liability. This idea of disjointed classifications is a well-known property of a well-constructed taxonomy, but it isn't commonly a machine-readable property in a taxonomy. The OWL data encoding language for ontologies provides standards for expressing this idea of disjointed classifications, equivalence classifications, pairwise classifications, and more.[18] A computer program, in this case an OWL reasoning system, may be used to validate a set of data—such as a taxonomy—as consistent with an ontology. The widespread use of taxonomies for disclosure modernization have been extremely useful because reporting entities can validate their data as conforming to the data description in a taxonomy. However, experience has shown that reporting entities sometimes construct taxonomies of poor quality,[19] causing the data to be of poor quality too. It is not effective to follow rules that themselves are not effective. Creating an ontology to specify the desired qualities of a taxonomy allows for the automated validation of taxonomy quality, which in turn brings about higher data quality.

These levels of agreement (L0 through L5) have existed in practice for centuries or more. Data encoding languages that represent these levels of agreement as machine-readable information have existed as non-proprietary global technology standards for a decade or more. Computer systems that use them also are commonly available to builders of data systems. No new technology needs to be invented.

Disclosure Modernization and U.S. Government Activities

In the U.S. government's 15-year history of disclosure modernization, three terms have informed federal requirements for disclosure modernization efforts: **machine-readable, open data, and disclosure modernization.**

As mentioned earlier, "machine-readable" means readable by computer software. The OPEN Government Data Act of 2018 (Title II of the Evidence Act) goes further, defining "machine-readable" data as "data in a format that can be easily processed by a computer without human intervention while ensuring no semantic meaning is lost."[20] The phrase "no semantic meaning is lost" means the data encoding language used must have sufficient resolution to capture meaning. This requirement for resolution drives us towards L4 and L5 (taxonomy and ontology) to ensure that "no semantic meaning is lost." Yes, a reader of a report may be able to look up a label to discover its meaning but that involves a human being, whereas our goal is to enable a computer to operate without human intervention. The distinct reference to "semantic meaning" also eliminates from consideration e-documents because, while they offer significant resolution, that resolution represents page layout rather than semantics. At best, e-documents provide for retrieval based on string searches only.

Open data is machine-readable data whose syntax is non-proprietary i.e., a non-proprietary technical standard, available for use without fees. Two critical points for consideration by policy-makers are (1) if a standard is actively managed for the benefit of users of the standard, and (2) if fees are charged for users of the standard. Some standards are proprietary, such as the data universal numbering system (DUNS) and also Westlaw citations. Those standards are controlled by their for-profit owners, and some users are charged fees for use of those standards. In the cases of DUNS and Westlaw citations, several government agencies have mandated their use, and thus their owners have a government-mandated monopoly. For the eight year period of 2010-2018, the GSA paid more than $131 million in fees to use DUNS.[21] With open data alternatives to DUNS,[22] this money could be put to better use.

Consensus standards bodies manage and promulgate non-proprietary standards. The World Wide Web Consortium (W3C) is a prominent example of a consensus standards body. The W3C maintains most of the standards associated with the world wide web including the globally popular data standards of HTML, XML, XLINK, and CSS. The XBRL standard is an open data standard widely used in disclosure modernization efforts globally. The XBRL standard is built on several W3C standards (e.g., XML and XLINK), and is maintained by the global not-for-profit consortium, XBRL International Inc.

Disclosure modernization is a movement of disclosure and reporting practices from e-document formats to open data. The goal of disclosure modernization is the exchange of information. And while the Portable Document Format (PDF) does provide for information exchange between persons or between a document editor and printing devices, it is a poor mechanism for exchanging useful data between computer systems. The description of page layout, which is the focus of the PDF, is not suited to working with the document's contents as semantic data, and therefore PDF is ill suited to automate data aggregation, validation, or analysis.

Government policies and practices are consistently turning in the direction of disclosure modernization. Some federal agencies already have or are actively transitioning to open data for compliance reporting to the government, including the FDIC (in 2005), the SEC (in 2009), and FERC (in 2021). Recent federal laws are bringing this about as a matter of multi-agency policy practice: the Digital Accountability and Transparency Act of 2014 (DATA Act),[23] the Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act),[24] and the Grant Reporting Efficiency and Agreements Transparency Act of 2019 (GREAT Act).[25]

Reaching further back to 2005 is the disclosure modernization work of the **Federal Deposit Insurance Corporation** (FDIC) and the **Federal Financial Institutions Examination Council** (FFIEC), the latter's purpose is to promote "uniformity and consistency in the supervision of financial institutions."[26] The FDIC, one of the members of the Council, implemented the FFIEC's requirement that banks provide quarterly Reports of Condition and Income (Call Reports) in the XBRL open data format. By doing so, the FDIC replaced a form previously submitted via paper and PDF, collecting approximately 1200 financial statement items. By 2005, about 8000 banks were filing an open data version of their Call Report.[27] The benefits quickly accrued: 100% of data submissions met FFIEC mathematical validation requirements versus 70% in the legacy system; data was made available to agencies within one hour whereas the legacy system required several days; and data availability to the public immediately followed calendar quarter end whereas the legacy system required several weeks.[28] Of further benefit, analysts completed their quarterly reviews seven days faster than they did with the legacy system. The efficiency increase enabled them to increase their caseloads by 20%.[29]

The **Securities and Exchange Commission** (SEC) in 2009 rolled out a mandate for public company reporting in the XBRL open data language.[30] Since that time, nearly every public company in the U.S. has reported their quarterly and annual reports to the SEC as machine-readable data. The SEC, through the Financial Accounting Standards Board (FASB) maintains an annually-updated taxonomy (L4) that public companies use to create their machine-readable reports. The SEC provides a near-real time RSS feed of these submissions. These are used by data aggregators such as Bloomberg and Morningstar, by institutional and retail investors, and by industry institutions. The CFA Institute recently demonstrated the value of this data with their analysis of FASB proposed changes to accounting policies in goodwill accounting.[31] The XBRL data from the SEC was used to evaluate that proposal subsequently to show that the change would result in "the write-off (amortization) over ten years of $5.6 trillion of assets on the books of U.S. public companies."[32] Data standards, especially at the more expressive levels of taxonomies (L4) and ontologies (L5), create demonstrably powerful opportunities for analysis.

In 2019, the **Federal Energy Regulatory Commission** (FERC) adopted a final rule to modernize certain annual disclosures by natural gas and electric utilities.[33] FERC announced its transition to standardized, machine-readable data to make it easier for companies to submit data and to enable FERC and other market participants to provide a faster, more accurate analysis of this energy company data. FERC also believes that, over time, this change will lower costs associated with preparing and filing compliance reports. Of great importance for their oversight responsibilities,

FERC expects efficiency improvements similar to those experienced by the FDIC: to process and analyze the data more quickly and correctly, and optimize caseload management. FERC expects their machine-readable program to be operational in 2021.

These are highlights of disclosure modernization efforts at the federal level. Many others exist at the agency level,[34] such as disclosure modernization for asset-backed securities disclosure and registration,[35] money market fund reform, amendments to Form PF,[36] crowdfunding,[37] and the listing standards for recovery of erroneously awarded compensation.[38]

These agency-specific modernization efforts are complemented by broader efforts that affect multiple agencies. The DATA Act required the U.S. Department of the Treasury to "establish Government-wide data standards for financial data and provide consistent, reliable, and searchable Government-wide spending data that is displayed accurately for taxpayers and policy makers."[39] The Treasury completed implementation on time (May 2017) and within budget.

Additional efforts also are underway across agencies through the Foundations for Evidence-Based Policymaking Act of 2018[40] and the Grant Reporting Efficiency and Agreements Transparency Act of 2019[41] (the GREAT Act). In fact, disclosure modernization efforts are well under way in every country of the European Union and many other countries around the world. This global effort is well past being a trend.

> **The U.S. government should see data policy as public policy.**

National, state, and local governments are major producers and consumers of data. Of the total $4.1 trillion outlay of the U.S. Federal Government in FY 2018, $2.25 billion[42] was spent on the 13 principal statistical agencies.[43] The need for data is clear. Therefore, we believe it is proper and beneficial public policy to promote disclosure modernization in order to maximize the value that may be extracted from this data.

The Data (R)evolution is Machine-Readable

The concepts of data and data standards are not new. The history of data parallels the development of science, and with science the development of complex societies. As social creatures it seems obvious we would establish socialization of our data, too, speedily sharing the benefits of experience and knowledge. In this collaboration both social and scientific, standards matter, and over time scientific advances were accompanied by advances in data practices.

The Scientific Revolution was a Data Revolution

Aristotle, as our first natural philosopher (4th Century BC), engaged in the scientific study of life. More than just noting observations he recorded them in structured form. In reading an excerpt of his observations one can see the structured, logical thinking in his prose.

> **Not all animals have a neck, but only those with the parts for the sake of which the neck is naturally present—these are the windpipe and the part known as the esophagus. Now the larynx is present by nature for the sake of breathing; for it is through this part that animals draw in and expel air when they inhale and exhale. This is why those without a lung have no neck, e.g. the kind consisting of the fish.**[44]

Readers can imagine how this prose might be the basis for a table detailing characteristics of animals, with discrete columns and conclusions reached by the analysis of the values in those columns.

ANIMAL NAME	HAS NECK	HAS LAYRNX	RESPIRATION THROUGH AIR
patagonia toothfish	false	false	false
angel shark	false	false	false
canine	true	true	true
feline	true	true	true

Illustration 9: Aristotle's observations

Aristotle developed ontologies for the natural world, achieving two thousand years ago the level of agreement that this paper today advocates for as L5. The scientific revolution of the 17th Century raised more stringent methods for validating obser-

vations and conclusions, and added expectations about the reproducibility of the results. Reproducibility drives agreement as to the meaning of data. This is why the scientific revolution also was a data revolution.

The Industrial Revolution was a Data Standards Revolution

The industrial revolution was characterized by production moving from assembly by hand and individual craft to assembly largely by machine and process. An assembly line with its interchangeable parts requires agreement on inputs and outputs.[45] Agreement is what makes parts interchangeable and thus a conceptual asset essential to the industrial revolution.

The development of the rail system provides a direct analogy to data. There was a time when railroad gauges were not standardized. When it came time to transfer goods from one train system to another, the cargo had to be transferred because the rail cars with incompatible gauges could not. Is this not similar to cutting and pasting between documents because the document formats are proprietary?

Containerized shipping, first practiced in 1956,[46] advances this data metaphor. Container standards enabled the same containers to be moved by ship and train and truck. The next advancement was to standardize the means of discovering the content of a container without unpacking. This was achieved by attaching RFID tags to the contents, making containers self-describing. This self-identification and self-description are conceptual assets essential to disclosure modernization.

The Internet Revolution was a Data Sharing Revolution

The Internet is, perhaps, the most visible example of data standards. It is the accumulation of various technical standards managed by domain-specific voluntary consensus standards bodies including the W3C, the Object Management Group (OMG), and the ISO. The movement to open source software provides standards for processing data. Contributions to the commonwealth of open source software comes from for-profit companies who recognize that such contributions 'raise all boats' including theirs. Consider the data challenges if there were proprietary languages for describing the content of a web page. Every web page would have to be reworked in each proprietary web page language and tested against each proprietary browser. Thankfully, it didn't play out in a proprietary model; the World Wide Web defeated proprietary on-line systems such as America Online (AOL) and Prodigy.

We now arrive at the currently active revolution, often called the 'fourth industrial revolution' where the data, in some sense, has left people behind and is able to collaborate independently of us.

Amidst a Data Intelligence Revolution

The data intelligence revolution began when people figured out systems and methods for mining data to extract valuable patterns. This domain of systems and methods is called machine learning (ML), which is part of the larger computer science domain of artificial intelligence (AI). In the past decade, AI has become a power tool, enabling computers to do tasks previously seen as the exclusive domain of human persons or even beyond the scope of both persons and hand-coded computer programs. As a tool, AI is something that changes the scale of one's capabilities, just as a bicycle changes the scale of distance—i.e., biking 5-miles takes less effort than walking. Machine learning changes the scale of pattern recognition, a fundamental tool of human intelligence that enhances our ability to recognize, classify, store, and use information effectively. With machine learning, we can learn from more examples than is possible for an individual person to review, and we may see patterns that would otherwise be invisible to the person.

Let's take an example based on optical character recognition. With machine learning a computer can learn to classify a set of pixels as a particular alphabetic character. It can then apply these learned classifications to new samples of pixel patterns to predict the alphabetic character represented by each sample. The example below illustrates the relationship between observation and classification.[47]

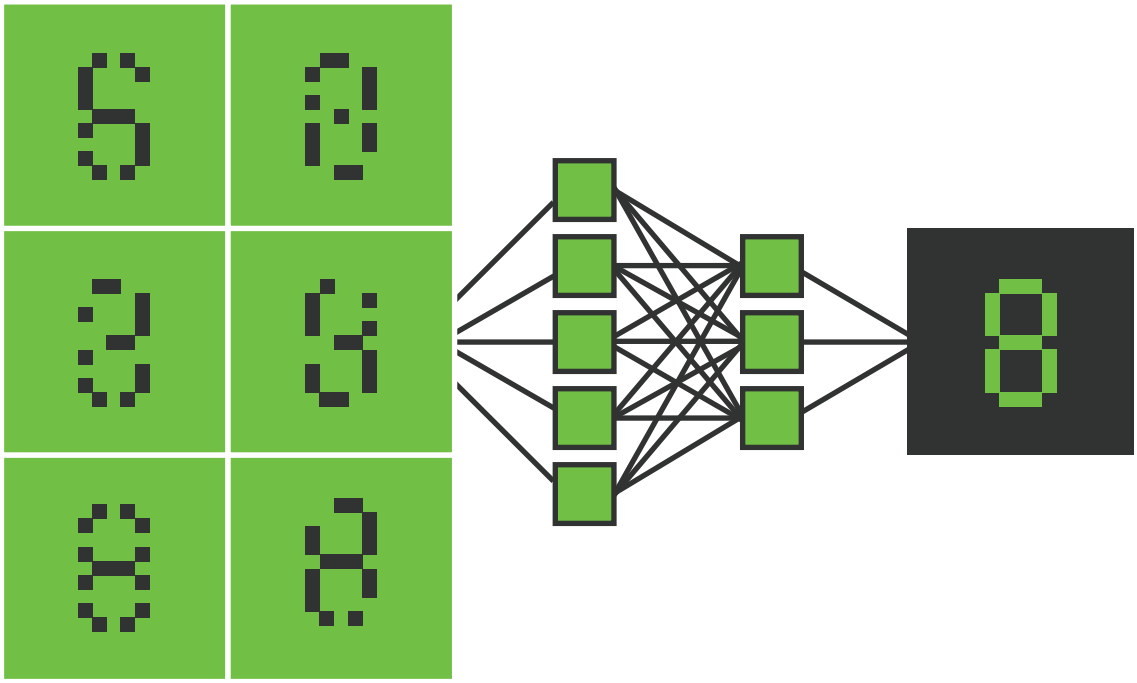


Illustration 10: Optical character recognition of patterns

Through an iterative self-training process, machines discover hypothetical patterns in data sets, and then apply this learning to detect these patterns in new data sets. The domain of reading an MRI similarly requires training, and radiologists have received special training to learn already-known patterns and then recognize these patterns in new MRI scans of patients. Machine learning changes the scale for the value extracted from an MRI because the machine's training can be done over millions of scans and detect patterns at the pixel level—both of these are at a scale

that cannot be performed by humans. This is similar optical character recognition which can detect characters in a bitmap—even a noisy bitmap. The idea of patterns is crucial to the data intelligence revolution.

> **It is the power of patterns that we seek to leverage when we move to higher levels of agreement of a data model because a model is just a structure, a pattern.**

Whether it is reading MRI scans, programming self-driving cars, or compliance reporting, the potential value may be enhanced by the qualitative understanding of the data model—the data patterns—and the agreement we have on those models.

It is timely to ask, Why can't ML be used for compliance reporting? Why bother with creating an explicit model of a compliance report when we have the option to let a computer 'machine learn' the model? To get to an answer, let's compare the starting points in each of the noted examples: MRI reading, self-driving cars, and compliance reporting. The model of an MRI scan is very simple; it is just bits comprising a visual image; it is a model based on a single kind of thing or element: the pixel. Since the model is relatively simple, ML can efficiently learn to detect patterns and anomalies.

Software for a self-driving car is different because it has to learn to navigate a seemingly much more complex model with more variables. Using a generated-from-scratch ML model for self-driving cars would require an immense amount of power with limited payoff. In this case, it is much more efficient to begin with an existing human-designed model that provides significant scope and precision. Specifically, the existing human-designed model provides more scope via multiple elements (e.g., transportation surfaces, traffic signals) and more precision via multiple attributes about each type of element (e.g., a traffic surface's attributes may represent speed limit, surface material and thickness, vehicle weight limit, etc.). In this case, software doesn't need to use ML to infer elements and attributes from a blank slate because a useful model already exists. Using this human-described model is much more effective and efficient than trying to ask the machine to infer all of these patterns from scratch.

The human standards help in another way. Programming a self-driving car should distinguish between common practice and prescribed practice. Consider the common practice that most but not all cars stop at red lights and stop signs. The prescribed practice is that all cars stop at red lights and stop signs. Surely, we want our self-driving cars to adhere to the prescribed practice while at the same time taking common practice into account when predicting the behavior of other cars.

The analogy between self-driving cars and compliance reporting is useful here. We have an existing human-designed model for compliance reporting, and that model has value because it accurately describes an idealized compliance

reporting system. We also want compliance reporting to follow prescribed practices for data entry rather than the common practice which may contain errors or fraud. Therefore, instead of having software learn from millions of data sets—some of which contain errors and fraud—we want to encode the human-designed model into the data standards so that machines are learning from the ideal set of standards and can distinguish them from reports which deviate from the ideal standards. We want to detect financial malfeasance (e.g., see Enron, and MCI Worldcom) not emulate it.

Compliance reports, particularly financial reports based on GAAP standards, are well-described logical systems that describe practices worth emulating. Encoding these logical systems into a machine-learning system allows the computer to use these human-derived standards to arrive at conclusions about compliance data. The encoding of logical relationships is what allows for inferences to be made, for reasoning to be applied. Returning to the accounting equation Assets = Liabilities + Equity, what we have is a logical assertion, i.e., that a value reported for Assets should equal the sum of the values reported for Liabilities and Equity. This logical relationship defines a pattern which can then be used by machine learning to detect situations where the assertions of fact do not comply.

Section Conclusion

In 1892, philosopher and psychologist William James wrote, "In a system, every fact is connected with every other by some thought-relation. The consequence is that every fact is retained by the combined suggestive power of all the other facts in the system."[48] The development of human knowledge is the practice of making connections. It may seem a leap to relate philosophy and psychology to compliance reporting, yet it is a pragmatic leap. Disclosure modernization should make that leap, too, to connect people and organizations with compliance data and compliance reports.

With this in mind, we recommend compliance standards as well as compliance data be machine-readable so these connections can be made.

> **Compliance reporting is ready to participate in the data intelligence revolution. What's needed is a policy push to achieve the level of agreement on data models necessary to realize it.**

Conclusion

Policy Recommendations

Three core recommendations extend naturally from the framework and information presented in this paper.

RECOMMENDATION #1:

Policymakers should require machine-readability when possible. Public policy for disclosure modernization should require both the reported compliance values and the compliance standards to be machine-readable. Machine-readable compliance standards enable technical innovations for automated reporting and data validation to reduce the compliance burden. Encoding domain knowledge as machine-readable data measurably improves the opportunities for analysis by both conventional and machine-learning methods, reducing the cost of governmental and stakeholder oversight while also increasing its effectiveness. Policymakers could implement this recommendation by incorporating language in proposed bills, regulations, and even guidance documents requiring machine-readability, similar to the standard applied in the OPEN Government Data Act.

RECOMMENDATION #2:

Policymakers should clearly communicate intent in legislative and regulatory actions on the role, purpose, scope of detail rendered as data, and applicability of data standards. The second recommendation is a necessary technical complement to the first. We recommend that federal data policy pursue disclosure modernization where data standards are expressed as L4 taxonomies or L5 ontologies. The transition to disclosure modernization recognizes modernization as a necessary commitment by the regulated and the regulators to ensure the integrity of compliance, evidence-based policymaking, and capital markets. Modern data practices reflect the understanding that all data are not created equal. More sophisticated demands upon data require more expressive levels of agreement for data standards. Modern data policy should communicate intent by describing the levels of agreement for data that should be met by those implementing policy. For compliance reporting, the very real necessity for data effectiveness (as noted in Recommendation #1) is met only by the more expressive levels of agreement on data standards: L4 taxonomies and L5 ontologies. Legislators can specifically support implementation of this recommendation by including expectations and purposes for data standards when drafting legislation, through bill text or committee reports. Similarly, regulators can better support this recommendation by clearly establishing expectations for standard effectiveness through proposed regulatory actions and guidance documents.

RECOMMENDATION #3:

Policymakers should encourage the adoption and use of open, consensus standards to encourage cooperation, efficiency, and innovation when drafting new data policies. The third recommendation is that technology choices should mini-

mize the technical and intellectual property obstacles to sharing and aggregating data. The goal is to enable agencies and data users—both public and private—to combine and analyze data (within statutory limits). The earlier example of the DUNS number is instructive as to the obstacles to effectiveness and the significant financial costs that proprietary standards bring. Open source software and mature data encoding standards that are free to use without restrictions and maintained by voluntary consensus bodies are multipliers of innovation. Governments should follow private enterprise in realizing the benefits of this technical and social cooperation. Policymakers can specifically incorporate these approaches by explicitly recognizing in legislative and regulatory actions the availability and prioritization of existing open standards when implementing new directives and policies.

Closing

Legislators and executive agency policy making in this area should address the three recommendations in Section 4.1, establishing disclosure modernization not just as best practice but as the standard practice. We summarize the practical implications of these recommendations:

- **Government agencies and other data consumers will be able to automate validation of data quality, the completeness of a reported data set, and the consistency of that data set. Further, they will be able to maintain effectiveness even as data volume and velocity increase.**
- **The machine-readable data will be detailed, high quality, and timely, available for both conventional analysis and the innovative analysis achieved by machine-learning and artificial intelligence platforms.**
- **It will build confidence in compliance and financial systems because with this data, fraud is caught more readily and earlier.**
- **People will be able to understand who gets funding and when, and how that funding is being utilized. This will improve government accountability and the performance of programs.**
- **Audits and data analysis will be democratized because vetted data is made available earlier, with greater scope, precision, and completeness.**
- **The data is a shared resource for citizens. This transparency encourages citizens to trust and engage their government. This trust is essential when a nation faces major events and crises (e.g., a pandemic, a natural disaster, a man-made disruption to infrastructure and/or society).**
- **Policymakers should recognize that disclosure modernization as described here supports the legitimacy and the proper functioning of compliance and financial systems, and most importantly, of governments themselves.**

Appendix A

Detailed Example

Below is a more thorough dissection of the example report from Section 2.2 and contains all of the observations from that section. Four sequenced illustrations capture the observations as seen through the eyes of a subject matter expert (SME) reading the Net Position summary of the Statement of Net Position. Illustrations 11 and 12 highlight the assertions of fact i.e., the values, and the classification(s) of these values. Illustrations 13 and 14 highlight the assertions of domain knowledge, i.e., the knowledge that experts bring to the example. Collectively, this is the knowledge-base we want to capture in a machine-readable form.

	A	B	C	D	E
1	**STATE OF GEORGIA** 5				
2	Statement of Net Position				
3	June 30, 2019 4	**Primary Government**			
4	(dollars in thousands) 3	**Governmental Activities**	**Business-type Activities**	**Total**	**Component Units**
5	**Net Position**				
6	Net Investment in Capital Assets (1)	20,361,680	8,429,136	25,566,212	3,534,685
7	**Restricted for:**				
8	Bond Covenants/Debt Service	64,016	–	64,016	94,454
9	Capital Projects	–	13,076	13,076	217,230
10	Guaranteed Revenue Debt Common Reserve Fund	53,766	–	53,766	
11	Loan and Grant Programs	–	–	–	1,844,780
12	Lottery for Education	1,354,630	–	1,354,630	–
13	Motor Fuel Tax Funds 1	3,508,961	–	3,508,961	–
14	**Nonexpendable:**				
15	Permanent Trust	–	181,016	181,016	2,398,751
16	Other Programs	–	–	–	46,977
17	Other Benefits	–	305,877	305,877	–
18	Other Purposes	1,293,746	313,732	1,607,478	407,868
19	Permanent Trust Expendable	–	–	–	804,556
20	Unemployment Compensation Benefits	–	2,535,856	2,535,856	–
21	**Unrestricted (1)**	(7,660,565)	(6,201,340)	(10,637,301)	2,295,805
22		$ 18,976,234	2 $ 5,577,353	$ 24,553,587	$ 11,645,106
23	**(1) Refer to Note 4 for additional details**				

Illustration 11

We start with the first part of assertions of fact: the values.

1 **There are 59 values in this section of the report.**

2 **Each is reported in US dollars (though only 4 are labeled as such).**

3 Each is reported in thousands, and we infer them to be rounded to the nearest thousand.

4 Each is measured at the end of the fiscal period, June 30, 2019.

5 Each is for the government entity called "State of Georgia."

The next illustration adds the classification(s) of each value, all of which are authoritatively defined by the Governmental Accounting Standards Board (GASB); local jurisdictions may provide further classifications within the GASB guidelines.[49] The assertions of fact consist of the values and their classification(s). Classifications assert the 'what' that is being measured by the values. For example, cell D9 contains the value 13076. Given the visual cues of the table, it is classified as the US dollar amount in thousands that is (a) "Restricted" to (b) "Capital Projects" of a (c) "Business-type Activity" of the (d) "Primary Government." As you may see from this example, a single value may have more than one classification. In these cases, it is classified as the aggregate of individual classifications.

7

	A	B	C	D	E
1	**STATE OF GEORGIA**				
2	Statement of Net Position				
3	June 30, 2019	**Primary Government**			
4	(dollars in thousands)	**Governmental 9 Activities**	**Business-type 10 Activities**	**11 Total**	**Component 12 Units**
5	**Net Position 6**				
6	Net Investment in Capital Assets (1)	20,361,680	8,429,136	25,566,212	3,534,685
7	**Restricted for: 13**				
8	Bond Covenants/Debt Service	64,016	–	64,016	94,454
9	Capital Projects	–	13,076	13,076	217,230
10	Guaranteed Revenue Debt Common Reserve Fund	53,766	–	53,766	
11	Loan and Grant Programs	–	–	–	1,844,780
12	Lottery for Education	1,354,630	–	1,354,630	–
13	Motor Fuel Tax Funds	3,508,961	–	3,508,961	–
14	**Nonexpendable: 15**				
15	Permanent Trust	–	181,016	181,016	2,398,751
16	Other Programs	–	–	–	46,977
17	Other Benefits	–	305,877	305,877	–
18	Other Purposes	1,293,746	313,732	1,607,478	407,868
19	Permanent Trust Expendable	–	–	–	804,556
20	Unemployment Compensation Benefits	–	2,535,856	2,535,856	–
21	**Unrestricted (1) 14**	(7,660,565)	(6,201,340)	(10,637,301)	2,295,805
22	**8**	**$ 18,976,234**	**$ 5,577,353**	**$ 24,553,587**	**$ 11,645,106**
23	**(1) Refer to Note 4 for additional details**				

Illustration 12

6 **All 59 values are classified as net positions.**

7 **Further classification of net positions are found in cells A6, A8:A13, A15:A20, and A21, for the values on each label's respective row.**

8 **The values in B22, C22, D22 and E22 are not explicitly classified. A SME would assume, because of the bar above and the double bars below the value, that this is the total net position for each of columns B, C, D, and E.**

9 **The values in column B are classified as "Governmental Activities."**

10 **The values in column C are classified as "Business-type Activities."**

11 **The values in column D are classified as "Total Primary Government."**

12 **The values in column E are classified as total "Component Units."**

13 **The net position classifications in cells A6 through A20 are further classified as restricted and expendable, or restricted and nonexpendable. From the point of view of a data consumer, it may be useful to think of expendable/nonexpendable as a classification of a classification because they apply only to net positions that are restricted. By having expendable/nonexpendable be a classification of restricted, a user could query the report to discover which restrictions were expendable and which were not. This is a query to learn about the nature of a government's organization rather than the values reported.**

14 **This complements A13 by classifying unrestricted items of net position, and it reinforces the value in classifying the state of restricted/unrestricted as separate from the classification of the different net positions.**

15 **This classifies net positions that are nonexpendable e.g., a trust. A SME would know that a trust may have an expendable portion e.g., a scholarship trust whose principal is not expendable but whose earning on the principal is expendable. An assumption is made that all the other classifications of A6 (other than rows 15 and 16) are classified as expendable.**

Illustration 9 highlights some mathematical relationships that contribute to the domain knowledge.

	A	B	C	D	E
1	**STATE OF GEORGIA**				
2	Statement of Net Position				
3	June 30, 2019	**Primary Government**			
4	(dollars in thousands)	**Governmental Activities**	**Business-type Activities**	**Total** 17	**Component Units**
5	**Net Position**				
6	Net Investment in Capital Assets (1)	20,361,680	8,429,136	25,566,212	3,534,685
7	**Restricted for:**				
8	Bond Covenants/Debt Service	64,016	–	64,016	94,454
9	Capital Projects	–	13,076	13,076	217,230
10	Guaranteed Revenue Debt Common Reserve Fund	53,766	–	53,766	
11	Loan and Grant Programs	–	–	–	1,844,780
12	Lottery for Education	1,354,630	–	1,354,630	–
13	Motor Fuel Tax Funds	3,508,961	–	3,508,961	–
14	**Nonexpendable:**			18	
15	Permanent Trust	–	181,016	181,016	2,398,751
16	Other Programs	–	–	–	46,977
17	Other Benefits	–	305,877	305,877	–
18	Other Purposes	1,293,746	313,732	1,607,478	407,868
19	Permanent Trust Expendable	–	–	–	804,556
20	Unemployment Compensation Benefits	–	2,535,856	2,535,856	–
21	**Unrestricted (1)**	(7,660,565)	(6,201,340)	(10,637,301)	2,295,805
22	16	$ 18,976,234	$ 5,577,353	$ 24,553,587	$ 11,645,106
23	**(1) Refer to Note 4 for additional details**				

Illustration 13

The specific observable mathematics:

16 **The value of row 22 should be the sum of the values above it in each respective column (i.e., columns B, C, D, and E).**

17 **The value of column E should be the sum of the values reported for columns B and C in each respective row. That is, the value reported for Primary Government should be the sum of the value reported for Governmental Activities and Business-type Activities.**

18 **According to A17 above, D6 should equal the sum of B6 and D6; and the value of D21 should equal the sum of B21 and C21. They do not. The SME would notice that they 'don't add up' and would have to investigate further, following the trail that continues through the footnote at the bottom left of the table.**

Finally, illustration 10 highlights some of a SME's non-mathematical domain knowledge. It is presented as a checklist to reflect how a SME might mentally 'check off' each as they either construct this statement as a preparer, or when they review the statement as a reader. In fact, the Government Financial Officers' Association (GFOA) has published and maintains a CAFR checklist[50] that is more detailed than the items below.[51] Detailed or not, currently, checklists like these are performed manually

with eyeballs and writing utensils. Effective disclosure modernization would enable automation of this production and review, and by doing so, promulgate the domain knowledge for compliance, mitigate inconsistencies, and detect or discourage fraud.[52] As a preparer or reader, building compliance knowledge into the data is tremendously valuable.

	A	B	C	D	E
1	**STATE OF GEORGIA** 21				
2	Statement of Net Position 19				
3	June 30, 2019 22		**Primary Government**		
4	(dollars in thousands) 23	**Governmental Activities**	**Business-type Activities**	**Total**	**Component 24 Units**
5	**Net Position** 20				
6	Net Investment in Capital Assets (1)	20,361,680	8,429,136	25,566,212	3,534,685
7	**Restricted for:** 26				
8	Bond Covenants/Debt Service	64,016	–	64,016	94,454
9	Capital Projects	–	13,076	13,076	217,230
10	Guaranteed Revenue Debt Common Reserve Fund	53,766	–	53,766	
11	Loan and Grant Programs	–	–	–	1,844,780
12	Lottery for Education	1,354,630	–	1,354,630	–
13	Motor Fuel Tax Funds	3,508,961	–	3,508,961	–
14	**Nonexpendable:** 28 27				
15	Permanent Trust	–	181,016	181,016	2,398,751
16	Other Programs	–	–	–	46,977
17	Other Benefits	–	305,877	305,877	–
18	Other Purposes	1,293,746	313,732	1,607,478	407,868
19	Permanent Trust Expendable	–	–	–	804,556
20	Unemployment Compensation Benefits	–	2,535,856	2,535,856	–
21	**Unrestricted (1)**	(7,660,565)	(6,201,340)	(10,637,301)	2,295,805
22	25	$ 18,976,234	$ 5,577,353	$ 24,553,587	$ 11,645,106
23	**(1) Refer to Note 4 for additional details**				

Illustration 14

19 **Does the CAFR contain a Statement of Net Position (SNP)?**

20 **Does the SNP include all the necessary parts for that statement: assets, deferred outflows, liabilities, deferred inflows, and net position? For our example above, we represent just the net position, so clearly this is not a complete SNP.**

21 **Is the entity identified? Does it comply with the identification requirements? The identification scheme is important because, for example, 31 states have a "Washington" county, and 28 have a "Jefferson" county.**

22 **Are all the values reported for the same date? Does that date coincide with the end of the fiscal reporting period for the entity?**

23 **Does the SNP contain information for governmental activities, business-type activities, and then primary government?**

24 **Does the SNP contain information for all discretely presented component units?**

25 **Does the SNP provide a total net position for governmental activities, business-type activities, primary government, and the discretely presented component units?**

26 **Net position may be classified as (a) net investment in capital assets, (b) restricted net position, or (c) unrestricted net position.**

27 **Does total net position equal the sum of net investment in capital assets, total restricted net position (expendable and nonexpendable), and total unrestricted net position? This particular example does not provide subtotals.**

28 **Are the components of restricted net position identified as expendable or nonexpendable?**

Automated validation requires that this domain knowledge be machine-readable. Additional domain knowledge that is desirable to capture are the connections between these particular value assertions and those reported elsewhere in the CAFR. These connections can be used to validate consistency of the entire CAFR.

Appendix B

A Summary of Technical Assertions

A summary of this paper's technical assertions follows:

1. **A data standard represents agreement between a sender and a receiver regarding the transmission of data.**
2. **Data may contain assertions of fact and assertions of domain knowledge; together they comprise the knowledge-base.**
3. **The quality of a data standard determines the potential value of the data.**
4. **The three classic problems of communication are the technical problem, the semantic problem, and the effectiveness problem.**
5. **Scope, precision and completeness serve as criteria for evaluating the effectiveness of a semantic solution.**
6. **We can categorize data standards as expressing one or more levels of agreement.**
7. **These are the six levels of agreement, arranged in order from least to most expressive: agreement as symbol (L0), agreement as location (L1), agreement as label (L2), agreement as schema (L3), agreement as taxonomy (L4), and agreement as ontology (L5).**
8. **Picking the level of agreement that best matches the knowledge-base is key to making the data standard effective.**

Endnotes

1. A very small sample of the numerous compliance standards that apply in the U.S. are the Health Insurance Portability and Accountability Act (HIPAA), the Federal Information Security Management Act (FISMA), FedRAMP, Sarbanes-Oxley Compliance (SOX), the Securities Act of 1933, the Securities Exchange Act of 1934, General Data Protection Regulation (GDPR), etc. Compliance standards maintain the stability and predictability of systems. The same may be said of compliance standards as was said by Oliver Wendall Holmes Jr. about taxes: "[They] are the price we pay for civilization."

2. Shannon, C. E., & Weaver, W. (1949). The Mathematical Theory of Communication. University of Illinois Press.

3. The acronyms represent: eXtensible Business Reporting Language, eXtensible Markup Language, JavaScript Object Notation, and Resource Description Framework, respectively.

4. Good for you, if you already noticed this.

5. As one example, accounting standards often are represented as codifications to indicate the rigor of this domain knowledge.

6. The acronyms represent the Securities and Exchange Commission, the Federal Energy Regulatory Commission, the Governmental Accounting Standards Board, the Financial Accounting Standards Board, the Municipal Securities Rulemaking Board, the Environmental Protection Agency, and the Federal Deposit Insurance Corporation, respectively.

7. Also known as the open world assumption.

8. Foundations for Evidence-Based Policymaking Act of 2018, P.L. 115-435 (2019). https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf

9. Municipal Securities Rulemaking Board. (2019). Muni Facts. http://www.msrb.org/msrb1/pdfs/MSRB-Muni-Facts.pdf

10. Ang, A. and Green, R. (2011). Lowering Borrowing Costs for States and Municipalities through CommonMuni. The Hamilton Project, Brookings Institution.

11. State Accounting Office, State of Georgia. (2019). Comprehensive Annual Financial Report Fiscal Year Ended June 30, 2019, p. 31. https://sao.georgia.gov/statewide-reporting/comprehensive-annual-financial-reports

12. The identification scheme is important because, for example, 31 states have a "Washington" county, and 28 have a "Jefferson" county.

13. Blackburn, D. (2018). Comprehensive Annual Financial Report, County of Will, IL. Will County Auditor. https://www.willcountyauditor.com/xbrl-cafr-2018

14. "Infon" is a specialized term for a unit of information.

15. We use the word "physical" in the logical sense of spatial location such as a spreadsheet coordinate. The true physical location, of course, is in computer storage.

16. These are the classifications defined by the Financial Accounting Standards Board (1985). Statement of financial accounting concepts no. 6: Elements of financial statements. https://www.fasb.org/pdf/aop_CON6.pdf

17. Gruber, T.R. (1993). Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies, 43(5-6), p. 908.

18. The Web Ontology Language (OWL) is a language standard managed by W3C. For a terse summary of its standard terms see OWL 2 Reference Card. (2009, October 18). W3C. http://www.w3.org/2007/OWL/refcard

19. Some reporting regimes permit reporting entities to modify the standard taxonomy (e.g., SEC public company reporting). The quality of these modifications vary. Other reporting regimes severely restrict taxonomy modifications (e.g., FERC).

20. Definitions, 44 U.S. Code § 3502. Government Publishing Office. https://www.govinfo.gov/app/details/USCODE-2011-title44/USCODE-2011-title44-chap35-subchapI-sec3502

21. Miller, J. (2017, February 21). GSA gives the 'dump DUNS' movement more hope. Federal News Network. https://federalnewsnetwork.com/reporters-notebook-jason-miller/2017/02/gsa-gives-dump-duns-movement-hope/

Endnotes (cont.)

22. One promising open data alternative is managed by the consensus standards body, Global Legal Entity Identifier Foundation. See Global Legal Entity Identifier Foundation. (2020). www.gleif.org

23. DATA Act, S.994 (2014). https://www.congress.gov/bill/113th-congress/senate-bill/994

24. Foundations for Evidence-Based Policymaking Act of 2018, P.L. 115-435 (2019). https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf

25. Grant Reporting Efficiency and Agreements Transparency Act of 2019, P.L. 116-103 (2019). https://www.congress.gov/116/plaws/publ103/PLAW-116publ103.pdf

26. FFIEC Home Page (2019). Federal Financial Institutions Examination Council. https://www.ffiec.gov/

27. Federal Financial Institutions Examination Council (March 30, 2006). FFIEC Annual Report 2005, p. 12. https://www.ffiec.gov/PDF/annrpt05.pdf

28. XBRL U.S.. FDIC Reporting. https://xbrl.us/home/filers/fdic-reporting/

29. Federal Financial Institutions Examination Council (2016, December). Improved business process through XBRL: A use case for business reporting, p. 6. https://xbrl.us/research/ffiec-2007/

30. See Securities and Exchange Commission (January 2, 2009). Final Rule: Interactive Data to Improve Financial Reporting, Release No. 33-9002. https://www.sec.gov/rules/final/2009/33-9002.pdf

31. FASB (2019, July 9). Invitation to comment: Identifiable intangible assets and subsequent accounting for goodwill. https://www.fasb.org/jsp/FASB/Document_C/DocumentPage?cid=1176172950529

32. CFA Institute (2020, January 13). Comment letter no. 103: Re: Invitation to comment (ITC), identifiable intangible assets and subsequent accounting for goodwill. https://www.fasb.org/cs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1175836097338&blobheader=application%2Fpdf&blobheadername2=Content-Length&blobheadername1=Content-Disposition&blobheadervalue2=1437883&blobheadervalue1=filename%3DINTANGGW.ITC.103.CFA_INSTITUTE_SANDRA_J._PETERS.pdf&blobcol=urldata&blobtable=MungoBlobs

33. Federal Energy Regulatory Commission (2019, June 20). Final Rule: Revisions to the Filing Process for Commission Forms, Docket No. RM19-12-000; Order No. 859. https://www.ferc.gov/whats-new/comm-meet/2019/062019/M-1.pdf

34. Joffe, M.D. (2018). Open data for financial reporting: Costs, benefits, and future. Data Foundation, pp. 35-36. https://www.datafoundation.org/xbrl-report-2017/

35. SEC Asset-Backed Securities Disclosure and Registration Rule, 79 FR 57183, Federal Register (2014). https://www.federalregister.gov/articles/2014/09/24/2014-21375/asset-backed-securities-disclosure-and-registration

36. SEC Money Market Fund Reform Rule; Amendments to Form PF, 79 FR 47735, Federal Register (2014). https://www.federalregister.gov/articles/2014/08/14/2014-17747/money-market-fund-reform-amendments-to-form-pf

37. SEC Crowdfunding Rule, 80 FR 71387, Federal Register (2015). https://www.federalregister.gov/articles/2015/11/16/2015-28220/crowdfunding

38. SEC Listing Standards for Recovery of Erroneously Awarded Compensation Release Nos. 33-9861, 34-75342, IC-31702; File No. S7-12-15 (proposed July 1, 2015). https://www.sec.gov/rules/proposed/2015/33-9861.pdf

39. Digital Accountability and Transparency Act of 2014, Public Law 113-101 (2014). https://www.congress.gov/113/plaws/pubL201/PLAW-113pubL201.pdf

40. Foundations for Evidence-Based Policymaking Act of 2018, P.L. 115-435 (2019). https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf

41. Maziarz, R. (2020, January 10). GREAT Act Becomes Law: What's Next for Grant Recipients. Data Coalition. https://www.datacoalition.org/great-act-becomes-law-whats-next-for-grant-recipients/

Endnotes (cont.)

42. Fatherree, K., & Hart, N. (2019). Funding the Evidence Act: Options for allocating resources to meet emerging data and evidence needs in the federal government. Data Foundation. https://www.datafoundation.org/funding-the-evidence-act-paper-2019

43. Other federal agencies as well as state and local governments spend additional money on data gathering, reporting, and analysis. Ironically, without disclosure modernization we lack a complete and accurate accounting of the total spent each year, and so have to estimate disclosure modernization's potential impact on this spending.

44. Lennox, J. (2019, Winter). Aristotle's Biology. In The Stanford Encyclopedia of Philosophy, Zalta, E.N. (ed.). https://plato.stanford.edu/archives/win2019/entries/aristotle-biology/

45. Cagle, K. (2019, August 23). The end of Agile. Forbes. https://www.forbes.com/sites/cognitiveworld/2019/08/23/the-end-of-agile/#5c2fb58d2071

46. Mayo, A. J., & Nohria, N (2005, October 2). The truck driver who reinvented shipping. Working Knowledge. https://hbswk.hbs.edu/item/the-truck-driver-who-reinvented-shipping

47. Illustration of recognizing the character "8" from a number of noisy samples is adapted from Cruz, J.A., & Wishart, D.S. (2006). Applications of machine learning in cancer prediction and prognosis. Cancer Informatics, 2, p. 63. https://journals.sagepub.com/doi/pdf/10.1177/117693510600200030

48. James, W. (1892). Psychology. Henry Holt and Company, p. 295.

49. Guidelines is a term of art meaning interpretation of standards.

50. Government Finance Officers Association. (2018). Certificate of achievement for excellence in financial reporting: General purpose checklist. https://www.gfoa.org/sites/default/files/GFOAGeneralPurposeChecklist.pdf

51. CAFR subject matter experts are encouraged to consider what is missing from the GFOA checklist and from the GASB codifications with regard to scope, precision, and completeness. These authorities, who maintain the domain knowledge for CAFRs, need to hear community requests for, and support of, CAFR disclosure modernization efforts.

52. Pope, K.R. (Director). (2015). All the queen's horses [Film]. United Kingdom: Gravitas Ventures. This documentary tells the story of how the comptroller in the city of Dixon, Illinois embezzled $53 million over a two decade period.

The Data Foundation is a non-profit organization that seeks to define an open future for our data, for a better government and society. The Data Foundation provides research, analysis, education, and programming that aims to improve the country's data policies. As the country's leading data policy think tank, the Data Foundation supports informed dialogue and thought leadership on emerging strategies and applications for accessing, using, and applying data in ways that are useful for decision-making.

http://www.datafoundation.org/

DATA FOUNDATION

Workiva

03

Standard Business Reporting:
Open Data to Cut Compliance Costs

17

DATA
FOUNDATION

pwc

DATA
FOUNDATION

The Data Foundation is the nation's first industry-focused open data research organization. Through research, education, and programming, the Foundation illuminates the benefits of transforming government information into standardized, open data. For more information, visit datafoundation.org.

pwc

PwC's Public Sector Practice helps federal agencies solve complex business issues, manage risk and add value through our comprehensive service offerings in financial management; program management; human capital; enterprise effectiveness; governance, risk and compliance; and technology, all of which are delivered seamlessly throughout the world. To find out more, visit www.pwc.com/publicsector.

Introduction

Regulatory compliance imposes heavy costs on the private sector. A 2014 study commissioned by the National Association of Manufacturers, for example, estimated that U.S. federal regulations cost companies $2.028 trillion annually.[1] A survey of manufacturing firms indicated that full-time staff and consultants devoted to regulatory compliance represented the majority of these costs.[2]

Regulatory compliance imposes heavy costs on government, too. In fiscal year 2016, the U.S. Securities and Exchange Commission estimated it would spend over half its budget to "foster and enforce compliance,"[3] the Federal Reserve Board of Governors' banking supervision and regulation division was its most expensive,[4] the Internal Revenue Service planned to invest over one-third of its budget on enforcement,[5] and the Census Bureau's most expensive program, aside from the five- and ten-year economic and population censuses themselves, was the one charged with maintaining its Business Register, with information on over 31 million U.S. business establishments.[6]

For the private sector, regulatory compliance involves compiling information and reporting it, at periodic intervals or when triggering events occur, to government agencies. For government, regulatory compliance involves receiving, reviewing, and acting on that information. For the private sector and government alike, these tasks involve a great deal of manual labor.

In all developed countries, including the United States, regulatory compliance is fragmented by industry and by purpose. Government agencies specialize in tax, securities, banking, statistics, workforce, environmental, and many other matters. Each agency, separately, has the legal authority to impose restrictions on, and collect information from, regulated companies and other entities.

Regulatory agencies' reporting requirements overlap with one another. For example, a 2011 study found that a large U.S. company was obliged to report substantially the same information, packaged differently, to the Securities and Exchange Commission, Federal Reserve, Census Bureau, and Bureau of Economic Analysis.[7]

Evidence demonstrates similar challenges faced by companies in the European Union. A consultation conducted by the European Commission Directorate General Financial Stability, Financial Services and Capital Markets Union, between September 2015 and January 2016 reveals that 288 respondents cited "Reporting and disclosure obligations" and "Overlaps, duplications and inconsistencies" as major hurdles.[8]

[A] 2011 study found that a large U.S. company was obliged to report substantially the same information, packaged differently, to the Securities and Exchange Commission, Federal Reserve, Census Bureau, and Bureau of Economic Analysis.

Around the world, governments are choosing to transform their information from disconnected documents into open data. For our purposes, the term *open data* refers to information that is made interoperable using standardized definitions and digital formats, and digitally *published* and freely available for use and reuse by its users.[9] The key, of course, is interoperability, which allows diverse systems and organizations to exchange and use one another's data without having to translate it.

For companies as well as agencies, open data offers significant efficiencies by reducing processing time and costs. First, if government agencies standardize data fields and formats for the information they collect, rather than expressing that information as unstructured documents, reporting companies' software can automatically compile and report it, reducing manual labor. Quality improves; human 'fat fingering' is eliminated.

Second, if multiple agencies align their fields and formats with one another by adopting universal standards for overlapping information, companies can submit the same information once, rather than multiple times to each agency.

Meanwhile, open data promises to cut regulatory agencies' costs and reduce their risks by allowing them to get and use regulatory information more quickly, shortening the processing required for data analysis. In the United States, for example, simple data matching could have revealed Bernie Madoff's fraudulent activities before his financial firm collapsed,[10] allowed agencies to quickly gauge the financial industry's exposure to Lehman Brothers while deciding whether to initiate a bailout,[11] and indicated that the fuel cell manufacturer Solyndra was the riskiest recipient of a federal loan guarantee well before its 2011 bankruptcy – if the relevant information had been available in a consumable format and in a timely manner. But because Madoff's securities reports, Lehman's financial filings, and Solyndra's energy and securities disclosures were available only as disconnected documents, not open data, these insights would have required expensive, time-consuming, and purpose-built analytics projects.

Most countries, including the United States, have not yet begun to apply open data to regulatory reporting. We will look at two prominent exceptions, the Netherlands and Australia, which have both embraced an approach known as Standard Business Reporting (SBR). SBR brings multiple government agencies together to define consistent data standards across their compliance requirements. In both the Netherlands and Australia, SBR reduces the manual labor of compliance, eliminates duplicated efforts of overlapping reporting requirements, and allows agencies to apply analytics.[12]

The Australian Tax Office estimated that Australia's SBR program saved the government and companies $1.1 billion in compliance costs during the 2015-16 fiscal year.[13]

With Australia and the Netherlands showing the way forward, this paper defines SBR, surveys the histories and results of the Dutch and Australian SBR programs, and envisions how a U.S. SBR program might begin, grow, and succeed.

Defining Standard Business Reporting

Standard Business Reporting (SBR) refers to the adoption of a common data structure across multiple regulatory agencies' reporting requirements. The path to SBR begins with creating that common structure, usually centered on a ***taxonomy***, or shared dictionary of data fields.

In an SBR environment, multiple agencies cooperate and agree to define the terms in the taxonomy and use it to represent the information that they collect from their regulated entities. Regulated entities then submit their regulatory information electronically, using the taxonomy. In some countries, including Australia, the SBR program expands to also include harmonized processes and channels of data submission through shared government centers.

In an SBR environment, companies can fulfill multiple regulatory reporting requirements simultaneously, while regulatory agencies can reduce or eliminate duplicative reporting systems. Meanwhile, more timely data allows agency investigators to instantly explore regulatory information across multiple reporting regimes. Finally, standardization opens new worlds of searchable data to investors and markets.

In a 2009 paper introducing the concept to tax agencies, the Organization for Economic Cooperation and Development (OECD) defined SBR using four main steps:

1. Creating a national … taxonomy which can be used by business to report [regulatory] information to Government. That taxonomy could encompass all [regulatory] data from outset or be built up gradually [;]
2. Using the creation of that taxonomy to drive out unnecessary or duplicated data descriptions [;]
3. Enabling use of that taxonomy for financial reporting to Government and facilitating straight-through reporting for many types of report[s] direct from accounting and reporting software in use by business and their intermediaries; and
4. Creating supporting mechanisms to make SBR efficient where they do not already exist (a single Government reporting service or portal or gateway, *etc.*)[.][14]

First, SBR involves the creation of an open data taxonomy that captures the information that private-sector companies must report to government agencies. In developed countries, regulatory agencies typically collect information from companies using document-based forms. In the United States, some agencies have modernized their forms by replacing documents with open data formats that convey each piece of information in its own defined field. Other agencies continue to collect forms as documents, in which all information is simply conveyed as unstructured text.[15]

Under SBR, however, these modernization efforts move forward in a more coordinated fashion, with multiple agencies agreeing on the same data elements. The ideal result is that multiple agencies' information collections

eventually become interoperable with one another: one open data set. The SBR approach focuses on common data elements and properties, rather than harmonizing processes or establishing shared data warehouses. Therefore, SBR brings standardization across multiple regulatory agencies while still honoring operational variations.

From a practical standpoint, implementing SBR requires top-down support through a single government entity empowered to create, impose, and maintain a taxonomy. Without a single agency or other authority designated as the lead entity, no one regulatory agency has sufficient authority or expertise to build a data structure expressing other agencies' reporting requirements.

Second, the taxonomy should identify duplicative reporting requirements and overlapping concepts, and either eliminate or consolidate them to bring all reporting requirements together into a single structure.

For example, where two agencies are collecting the same piece of information – a company's name, for instance – the taxonomy can reflect both with just one data field. Where two agencies are collecting substantially the same piece of information, the taxonomy might adopt a definition that can be used by both (or, in the case of mathematical relations, a business rule describing the mathematical difference). Or suppose the taxonomy's creators discover that two agencies are collecting similar pieces of information, but with some differences. This partial overlap is an opportunity to clarify the requirement and possibly eliminate it.

Third, companies must be required to comply with reporting requirements electronically, by submitting data encoded using the taxonomy. To this end, agencies participating in the SBR program must be prepared to accept reports electronically. Importantly, as in the case of the Dutch SBR program, the taxonomy may cover both requirements that are common across multiple agencies and those that are specific to just one agency.

Fourth, the lead entity should encourage the creation of supporting technologies, embracing public-private partnerships where necessary and practicable. These might include a single, government-wide electronic portal for regulatory reporting; platforms to integrate compliance with a company's other business functions; or private-sector software tools enabling users to access and analyze regulatory data.

The OECD points out that the third and fourth steps "are essentially technology infrastructure-related [steps] for which there are proven solutions."[16] The core challenges of the first and second steps, by contrast, are organizational and programmatic, not technological.

Standard Business Reporting in the Netherlands

The government of the Netherlands hosts the world's oldest and, to date, most successful SBR program. The program's roots date to 2002, when the Ministry of Economic Affairs initiated a public-private venture to use technological solutions to address the private sector's compliance burdens.[17] One of the solutions developed by this venture was a single, government-wide electronic address for the submission of regulatory filings, known as the *Overheidstransactiepoort*, or OTP.

Inspired by the OTP's concept of electronically-standardized reporting, the Ministries of Justice and Finance started the Netherlands Taxonomy Project (NTP) in 2004.[18] The NTP released the first version of its Netherlands Taxonomy, covering companies' financial and fiscal information, in June 2005. Despite efforts in the public and private sectors to move the concept forward, by the beginning of 2009, the Netherlands Taxonomy Project had not yet delivered the contemplated compliance savings. Fewer than 10,000 filings were being submitted annually using the Netherlands Taxonomy, out of the hundreds of thousands of regulatory submissions submitted each year by Dutch companies.[19]

The Dutch government concluded that the program was focusing on design over implementation, and refocused on implementing within just three agencies: the Chamber of Commerce, the Tax and Customs Administration, and Statistics Netherlands.[20] In particular, "the Tax and Customs Administration presented a plan to start phasing out (from 2013 onwards) the information exchange channel that competed with SBR."[21]

Meanwhile, the government adopted the term *Standard Business Reporting* to better communicate the program's

Dutch Standard Business Reporting

Business Regulators

Tax Administration

Semi-mandatory: SBR is the only way to file using a system-to-system connection. But there is an option to file manually using a web portal.

Chamber of Commerce

Semi-mandatory: SBR is the only way to file using a system-to-system connection. But there is an option to file manually using a web portal.

Education Executive Agency

Mandatory using a web-portal. System-to-system will be implemented in the near future.

Central Bureau of Statistics

Voluntary

Ministry of Internal Affairs

Pilot-phase

Private Sector

Banks

Semi-voluntary: When sending a credit report via paper, an enterprise must pay a fee. Using SBR is free.

Rembrandt Logistics NL

Hallo Technologieën

In the Netherlands, companies can comply with tax, business, and statistical reporting requirements automatically and electronically, using the same data. The Dutch SBR program includes voluntary, semi-mandatory, and mandatory participation by agencies.[22]

meaning and purpose. In the fall of 2009, the SBR Council, a new steering committee representing the highest level of the agencies involved, was set up, and responsibility for implementation was centralized within the government's digital service, now called Logius.[23]

Within two years nearly 100,000 annual regulatory filings – primarily tax and financial disclosures – were being submitted by companies to regulatory agencies using the taxonomy.

However, "[t]he government agencies understood that complete adoption by the market and the step to proper service management could not be realised if SBR remained a voluntary solution for business reporting."[24] Market participants acknowledged this as well.[25] Accordingly, in June 2011, the Tax and Customs Administration agreed to make the taxonomy semi-mandatory[26] for tax declarations by 2014.[27] Tax declarations submitted directly from software packages were required to be encoded in the Netherlands Taxonomy. Companies could still manually submit tax declarations using a Tax & Customs Administration web portal that automatically converted the information into the taxonomy.

In 2012, the Chamber of Commerce, which collects Dutch corporate financial disclosures, began to discourage companies from submitting financial statements in PDF and encourage them to submit financial statements using the taxonomy.[28] In November 2015, the Dutch parliament enacted legislation to make financial reporting via the Netherlands Taxonomy semi-mandatory, in the same manner used for tax declarations, for smaller companies in 2017, medium-sized companies in 2017, and larger companies in 2019.[29]

Beyond the domains of tax and corporate finance reporting, the Netherlands Taxonomy has been extended to cover additional regulatory reporting regimes. On September 8, 2015, Standard Business Commissioner Rob Kuipers laid out a vision for "business reporting in zero clicks": the idea that *all* regulatory reporting should be automatic, with information automatically culled from companies' software.[30]

The Netherlands Taxonomy has also been embraced beyond the regulatory sector for some business-to-business reporting practices. On January 1, 2017, the country's three largest banks finished converting their commercial credit reporting regimes to the Netherlands Taxonomy.[31]

The creation and development of the Netherlands Taxonomy is a model of the first and second steps of SBR, as defined by the OECD. "In the first stage of taxonomy development 200,000 reporting data items were identified. After thorough analysis the number was reduced to 8,000 unique data items."[32] The number of unique data elements has been further reduced to approximately 4,500 through legislative reforms.[33]

The Dutch government has figuratively and literally written the book on SBR. Challenging the Chain: Governing the Automated Exchange and Processing of Business Information, which describes the history, challenges and approaches of the SBR program, is freely available online.[34] The 'chain' of the title is an *information chain* – a collection of links of information connecting suppliers (such as regulated companies) and users (including regulatory agencies and other users). Each link represents an action on the data that contributes to the value of the process or deliverable, thus establishing a *flow*. Information chains are the specific focus of Challenging the Chain. The book provides a simple example of the concept:

> When considering the production of a consumer good, such as a television, activities could involve its assemblage, production, inspection, packaging, transport and storage. These kinds of activities require inputs - human resources and raw materials - which are employed to add value and transport the flow elements to the next actor (or stage) in the chain.[35]

It is easy to imagine the information chains related to regulatory reporting processes. Regulated entities accumulate, validate, and report prescribed information to regulatory agencies, who then consume the information for analysis, enforcement, and oversight, and sometimes make it available to the public.

By envisioning all of its regulatory reporting requirements as information chains, the Dutch government was able to develop a holistic view of its regulatory environment that is not dependent on any particular domain (tax, corporate finance, statistics, *etc.*).

Thanks to the Dutch SBR program, Dutch companies enjoy both benefits identified in the introduction to this paper. First, their software uses the Netherlands Taxonomy to automatically compile and submit tax declarations, financial statements, and other regulatory

filings. This saves time and money over manually preparing disclosure documents. Second, the taxonomy's reduction in data elements from hundreds of thousands to just a few thousand allows companies to invest time and money just once to comply with multiple agencies' requirements.

Meanwhile, by eliminating documents and PDFs from their intake, and replacing document-based reporting with open data, regulatory agencies in the Netherlands have gained the ability to deploy analytics without any translation.

Other stakeholders in the Netherlands recognized the value of open, electronically standardized data and have begun implementing the Netherlands Taxonomy for other purposes. For instance, ING Bank has reduced fees for small and medium companies, together with other benefits, if their financing applications include SBR-enabled digital financial information.[36]

Standard Business Reporting in Australia

On October 12, 2005, Prime Minister John Howard's Australian government appointed a task force "to identify practical options for alleviating the compliance burden on business from Government regulation."[37] The task force's resulting report,[38] issued on January 31, 2006, recommended that the government review all the "data collection and regulatory reporting obligations" imposed by its financial and economic agencies, with a view toward eliminating overlaps, and "establish an integrated data collection portal to ensure that regulated entities have to provide information only once."[39]

To implement this recommendation, the Australian government established a Standard Business Reporting program in 2008, involving both federal and state/territorial agencies, with the Australian Taxation Office (ATO) serving as the lead agency. While the initiative was ultimately projected to involve the "whole-of-government," consideration was given in the initial stages to the inclusion of two groups of agencies: the "General Ledger" cluster, which receive taxation, financial and company reports from businesses, and the "Trade" cluster, which receive customs and international trade reports.

Initially, the SBR program's leadership focused only on the "General Ledger" cluster of agencies. Following the publication of the first version of the Standard Business Reporting Australia Taxonomy in 2010,[40] the ATO, the Australian Securities & Investments Commission (ASIC), and the revenue offices of Australian states and territories began accepting filings using the taxonomy.[41] The Australian Prudential Regulation Authority, which regulates banks, retirement funds, credit unions, and other financial entities, began accepting SBR filings the following year.[42] The Australian Bureau of Statistics also participated in the program through the development of the taxonomy.

The number of software developers with SBR-enabled commercial software products fit for purpose and ready or available to be used by Australian companies steadily increased to 54 by mid-2014, covering 515 reporting obligations to the "General Ledger" agencies.[43] By this time, a further 115 developers were licensed to develop SBR-enabled software for businesses to compile and submit reports using the taxonomy.[44]

The Australian SBR program has differed from the Netherlands' program in that, with the sole exception of superannuation (retirement fund) transactions, which arose from a request by the industry, the use of SBR has not yet been mandated.

Like in the Netherlands, the Australian SBR program has begun to spread to business-to-business, as well as business-to-government, reporting. A pilot effort at the Australian Business Register (somewhat equivalent to U.S. corporate registers, which operate on the state, rather than federal, level) is aimed at standardizing electronic invoices issued by companies to one another, using new data elements in the same taxonomy.[45]

Also paralleling the Dutch experience, the creation of the Australian taxonomy reduced the number of data elements Australian companies report to the participating regulatory agencies. By 2014, "implementing the taxonomy … reduced the number of unique reporting terms used across the forms that [were] accessible through SBR from almost 35,000 to less than 7,000 unique terms, a reduction of more than 80%."[46]

To assist in expanding the scope of SBR to a wider

Australian Standard Business Reporting

Business Regulators

- ACT Revenue Office
- NSW Office of State Revenue
- NT Revenue Office
- QLD Office of State Revenue
- SA Office of State Revenue
- Tasmanian State Revenue Office
- Victorian State Revenue Office
- WA Office of State Revenue
- Australian Taxation Office
- Australian Securities and Investment Commission
- Australian Prudential Regulation Authority
- Australian Business Register
- GNGB Superannuation B2G
- Superannuation Funds B2B

Private Sector

- Electronic Invoicing

Pilot-phase

Koala-fied Truckers, Inc.

Aussie Pharma Corp.

In Australia, data standards enable companies' software to automate compliance tasks by pulling information from internal systems.[47]

"whole-of-government" context, the program's leadership decided in 2015 to support a limited number of data formats, in addition to XBRL (eXtensible Business Reporting Language), which had been used since the program's launch. The decision to broaden SBR's capabilities was made to support the implementation of different patterns of digital interaction in additional sectors of reporting to government.

"We stood on the shoulders of the Dutch efforts," says John McAlister, the Assistant Commissioner, Stakeholder Engagement and Adoption (ABR and SBR), at the Australian Business Register.[48] The results, as gathered from various Australian government sources, speak for themselves. In 2014-15, some AUS $400 million in compliance savings were recorded in Australia due to SBR across two spheres: business-to-business, and business-to-government.[49] An ATO/Deloitte report estimated savings of AUS $1.1 billion for 2015-16; these savings are projected to increase substantially in future years as the SBR program further expands.[50]

Australian companies are reaping the same benefits as Dutch ones: automated reporting and reduced regulatory overlap. The Australian government benefits in the same way as its Dutch role model: standardized filings facilitate analytics.

Standard Business Reporting in the United States

Despite the early success of SBR in the Netherlands and Australia, and efforts underway in other countries,[51] the concept is nearly unknown in the United States.[52] Admittedly, the U.S.' regulatory landscape is more complicated than the Netherlands' or Australia's, its economy many times larger, and compliance costs correspondingly enormous. A U.S. SBR program would face greater challenges than its Dutch and Australian predecessors – but could also expect to generate even greater savings for the private sector and government.

On several occasions, the U.S. private sector has tried to build open data taxonomies to standardize regulatory information.

- In March 2006, Stanford Law School's Rock Center for Corporate Governance announced the creation of the **Open Source Corporate Governance Reporting System**, a database of publicly-traded companies' corporate-governance data, which would "enable users to develop their own research and analytics."[53]
- In May 2009, the Depository Trust & Clearing Corporation, which manages clearing services for most U.S. financial products, and the standards organizations SWIFT and XBRL US announced[54] that it would build an open data taxonomy to express **corporate actions**, such as the issuance of dividends, rights offerings, and mergers. The taxonomy was completed in 2011 and updated in 2012, but is not in use.[55]
- In June 2016, State Street Bank completed the first successful pilot test of the **Financial Industry Business Ontology (FIBO)**, which standardizes financial-industry concepts.[56] FIBO has been in development by the Enterprise Data Management Council since 2011, but is not used in production systems or in official regulatory reporting.

None of these efforts has achieved market-wide savings in regulated entities' compliance costs. Nor have any, more broadly, made information acquisition more efficient for the industries in which they operate.

Since regulatory reporting requirements are driven by government, it makes sense that that any SBR initiative must be also led by government. This is not any easy task, given the cultural and political hurdles facing any change to regulatory practice. Nonetheless, there are significant economic benefits to be gained, along with the promise of better, more timely information.

How would the US Government start down an SBR path, pursuing the four steps identified by the OECD? First, which U.S. government entity is best positioned to create a national taxonomy?

The United States is a constitutional republic rather than a parliamentary democracy like the Netherlands and Australia, with separate legislative and executive branches. A U.S. government-wide taxonomy creation project would need a formal, legal mandate from

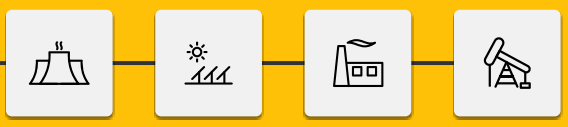
U.S. Standard Business Reporting

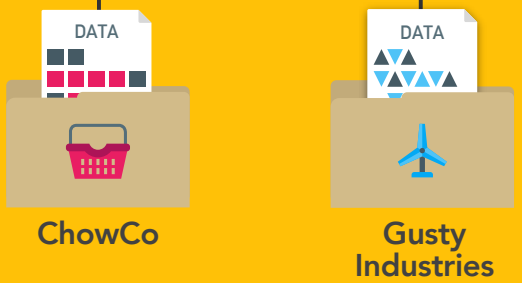
General Business Regulators

- Department of the Treasury
  - Internal Revenue Service
- Department of Labor
  - Employee Benefits Security Administration
  - Bureau of Labor Statistics
  - Occupational Safety and Health Administration
- Securities and Exchange Commission
- Federal Trade Commission
- Department of Commerce
  - Census Bureau
  - Bureau of Economic Analysis
- Business Registrars
- State Revenue Departments

AL AK AS AZ AR CA CO CT
DE DC FL GA HI ID IL IN
IA KS KY LA ME MD MA MI
MN MS MO MT NE NV NH NJ
NM NY NC ND MP OH OK OR
PA PR RI SC SD TN TX UT
VA WA WV WI WY

Industry-Specific Regulators





ChowCo

Gusty Industries

A U.S. Standard Business Reporting program would use data standards to consolidate and automate reporting to general and industry-specific regulators.

Congress – and preferably informal support from both Congress and the White House – in order to move forward across multiple regulatory sectors.

In both the Netherlands and Australia, tax reporting is a centerpiece of SBR, and tax agencies have played a key role in taxonomy creation. But the U.S. Internal Revenue Service has characteristics that quite distinguish it from its Dutch and Australian counterparts: an institutional and statutory reluctance to share data and a cultural aloofness from other regulators. Likewise, there are also very strong traditions of independence associated with U.S. banking and securities regulators.

Several existing entities and projects within the U.S. federal government have government-wide regulatory or data-related roles. Under the DATA Act of 2014, the Treasury Department has promulgated a government-wide taxonomy for U.S. federal spending information. The White House Office of Management and Budget's Office of Information and Regulatory Affairs (OIRA) reviews government information collections under the Paperwork Reduction Act;[57] the Financial Stability Oversight Council (FSOC) oversees all financial regulatory reporting, as mandated by the Dodd-Frank financial reform;[58] and the National Information Exchange Model is the only government-wide, multi-domain data standardization project.[59] However, none of these are currently capable of launching or sustaining a U.S. SBR program. The DATA Act mandate applies to spending information, not regulatory reports. OIRA focuses on document-based forms, not data, and its authority over independent, non-Cabinet agencies is limited; the FSOC's membership is restricted to financial regulatory agencies; and NIEM's expansion is voluntary, not mandatory.

Since no U.S. government entity currently has the authority to create and impose an SBR taxonomy, a full-scale SBR program cannot begin without a legislative reform, enacted by Congress, that either creates and empowers a new entity or else invests an existing one with new powers. Moreover, even with a legal mandate from Congress, such a complex project probably would still fail without enthusiastic support from the 'bully pulpit' of the Presidency.

While the idea of central authority is important, it may be equally important to not assume that there will be just one SBR taxonomy, at least not initially. For example, the FSOC already has the authority to create and impose a taxonomy within the financial regulatory sector.

Second, how could a U.S. SBR taxonomy be wielded to consolidate duplicative requirements? The entity leading a U.S. SBR program must be assigned to investigate, and publish reports on, duplication across regulatory agencies' reporting regimes revealed by the taxonomy. Agencies' own constituencies – whether securities or retirement or tax – must be encouraged to advocate for those agencies to change local rules and reporting practices to eliminate duplication. Without a powerful mandate for consolidation, domain-specific and agency-specific idiosyncrasies will rule the day.

Third, how could companies be required, or at least strongly encouraged, to file reports using a U.S. SBR taxonomy? Each agency participating in a U.S. SBR program could direct its regulated entities to do so – but a stronger centralized mandate might be required. Perhaps Congress should amend regulatory agencies' governing laws to require each agency to direct its regulated entities to submit regulatory information using the SBR taxonomy.

Moving regulated entities to report under the SBR taxonomy is critical to step four in the OCED recommendations, *i.e.*, having available the necessary tools for the compilation and submission of SBR reports, and for the analysis of standardized SBR data. The technology industry will not invest in creating the necessary tools without the expectation of a market, and phased-in mandatory reporting would provide such notice. Moreover, a U.S. SBR program cannot expect significant benefits without involving software providers, a lesson learned from the Dutch and Australian SBR experiences.

John Truzzolino, Director, Business Development for Global Capital Markets at RR Donnelley reinforced this necessity. "Through this process what is really the big differentiator ... what is challenging in the US ... [is that] the international markets are more open to presenting a true consortium of public and private stakeholders working in concert on a common objective. The Dutch program involved a covenant of cooperation [between government and industry]."[60]

Finally, are the U.S. government and private sector ready to embrace the challenge of SBR? One might assume that the immediate answer is 'YES!', especially considering the chaos and expense of the current regulatory compliance landscape, and recent related moves by the U.S. Congress and executive branch. But as noted earlier,

agencies are slow to embrace change, especially change which suggests more transparency and better oversight. These are powerful agencies with powerful constituencies, both rich in tradition and culture, and to embrace this sort of change will require extraordinary leadership and vision.

One way to test the appetite for SBR is to track current related legislation. The U.S Congress is considering the Financial Transparency Act,[61] which, if enacted as written currently, would require the eight major U.S. financial regulatory agencies to adopt standardized data fields and formats for the information they collect from public companies, banks, and financial firms. The Financial Transparency Act, if enacted, could have the effect of establishing a more modest version of SBR, limited to financial regulatory compliance regimes.

The U.S. government has embraced data standardization for its own internal financial reporting. Under the Digital Accountability and Transparency Act (DATA Act) of 2014,[62] by May 2017 every federal agency must begin reporting its spending using a standardized taxonomy created by the Treasury Department.

If the United States embraces SBR, we can expect benefits similar to those enjoyed already in the Netherlands and Australia. Companies will be able to automate the compilation and submission of their regulatory reports – and enjoy additional cost reduction from the ability to reuse efforts from one regulatory regime to another. Regulatory agencies will benefit from access to a standardized data source covering multiple regimes – a foundation for instant analytics to illuminate violations, errors, and fraud.

The Future of Standard Business Reporting

Assuming, for the moment, that the United States initiates and executes an SBR program along similar lines to the pioneering ones of the Netherlands and Australia, the first beneficiaries, as described above, will be regulated companies and regulatory agencies.

As Danny Kermode puts it, "I think everyone agrees interactive data is the future, they just don't understand how to get there."[63] Mr. Kermode is the Assistant Director of Water & Transportation for the Washington Utilities and Transportation Commission (UTC), and initiated a pilot program to adopt a standardized taxonomy for utilities' financial reports. The UTC had been collecting financial reports in multiple formats, requiring both a PDF file and also the native format (Word, Excel, etc.) that produced the file, but it realized advantages from adopting the eXtensible Business Reporting Language (XBRL) format, which expresses the arithmetic relationships between data fields.

"I see the benefits for regulators as enormous," said Kermode. "Recognizing that we have a lot of highly skilled professionals who spend a tremendous amount of time doing classic transcription[, standardized reporting] will allow these individuals to focus on analysis and spend time understanding the data, rather than transcribing." At the same time, he noted that companies would be able to communicate the data internally to any element of their company using the same setup and technology.

Aside from regulatory agencies and regulated companies, who else benefits? *Information Week* recently reported that IDC forecasts a 50% uptick in the sale of big data and business analytics software, hardware and services by 2019, with sales reaching $187 billion.[64] To the extent that SBR delivers an open data set covering information reported to multiple regulators, SBR will create new opportunities for:

- **FINANCIAL TECHNOLOGY FIRMS** - whether their goal is to work in concert with existing financial services firms or to disrupt them. As Kevin Roose wrote in *New York* magazine, "Undercutting big banks and speeding up processes might not be as sexy as, say, creating the next Snapchat, but it's low-hanging fruit for techies who want a way in to a lucrative market. After all, today's megabanks are really just bundles of particular, loosely related services cobbled together by years of acquisitions and market-consolidation. If those bundles can be broken apart, the start-up world's revolution looks a lot more plausible."[65]

- **EXISTING FINANCIAL FIRMS**, like Citigroup, Wells Fargo, BBVA and others, who have made a point of increasing their focus on venture funding and acquisition of firms like the ones above, or who have or are developing massive internal engines to handle transaction data, portfolio and/or risk management and even client services (institutional or otherwise), in such a way that they will have access to proprietary big data that can drive major strategic initiatives.

- **THIRD-PARTY SOFTWARE AND SERVICE VENDORS**, ranging from tax firms and auditors to compliance outsourcing or consulting firms such as 8of9 or RIA in a Box, all of which are accustomed to helping clients navigate through unfamiliar process and technology as an element of their business model.

- **MARKET DATA FIRMS**, who will no doubt be keeping a close eye on how new technologies - and the availability of extensive troves of public financial data - might affect them in the future, and who have the existing technology, expertise and branding to be associated with data as new models emerge.

SBR creates efficient, trusted, open data ecosystems – or information chains - involving regulatory agencies, regulated entities, and other organizations. Once standardized and trustworthy, these ecosystems (or chains) do not only serve existing stakeholders, but also contribute to the further generation of innovative information products and services.

SBR is no longer new. The necessary technologies and governance structures have been tested and refined by the Dutch and Australian programs. All that is needed in the United States is the will to seize this opportunity. To realize savings for regulated companies and the regulators themselves, and to enable new growth in financial technology and market data, Congress and the executive branch must act. ■

REFERENCES

1. National Association of Manufacturers, *The Cost of Federal Regulation to the U.S. Economy, Manufacturing and Small Business*, September 10, 2014, *available at* http://www.nam.org/Data-and-Reports/Reports/Cost-of-Federal-Regulations/The-Cost-of-Federal-Regulation/.

2. *Id. at 22*.

3. U.S. Securities and Exchange Commission, *FY 2017 Congressional Budget Justification, available at* www.sec.gov/about/reports/secfy17congbudgjust.pdf, at 16.

4. Board of Governors of the Federal Reserve System, *Annual Report 2015*, *available at* www.federalreserve.gov/publications/annual-report/2015-federal-reserve-system-budgets.htm - xboardofgovernorsbudgets-f2c579a2, at Table 3.

5. Internal Revenue Service, *FY 2017 President's Budget, available at* www.treasury.gov/about/budget-performance/CJ17/02-06.%20IRS%20FY%202017%20CJ%201%2022%2016%20v2%20FINAL%20CLEAN.PDF, at Table 1.1.

6. U.S. Census Bureau, *FY 2016 Budget Estimates, available at* www.osec.doc.gov/bmi/budget/FY16CJ/Census_2016_CJ.pdf, at 24 ($65 million for "General economic statistics").

7. XBRL US, *Better Data for Better Decisions: Standards to Improve Corporate Government Reporting*, October 2011, *available at* xbrl.us/wp-content/uploads/2011/12/BetterReporting.pdf, at 10 (reporting that a majority of data fields reported to the Bureau of Economic Analysis by a large public company were already reported to the Securities and Exchange Commission; the same was true for a substantial proportion of data fields reported to the Census Bureau; and Federal Reserve reports were entirely duplicative of SEC reports).

8. Peter van den Hul, *Financial Data Standardisation Project*, 22nf Eurofiling Workshop, June 2, 2016, *available at* http://eurofiling.info/201606/Presentations/2016-06-02%2015h00%20FDS%20Project%20Presentation.pdf.

9. Alison Gill, Adam Hughes, and Hudson Hollister, *The State of the Union of Open Data*, November 2016, *available at* www.datafoundation.org/state-of-the-union-of-open-data-2016/.

10. *See* Randy Hultgren, "How to Stop the Next Bernie Madoff," *The Guardian*, August 16, 2016, *available at* www.theguardian.com/commentisfree/2016/aug/16/how-to-stop-the-next-bernie-madoff.

11. *See* Dick Berner, *Remarks by OFR Director Dick Berner at the Financial Regulation Summit: Data Transparency Transformation*, March 24, 2015, *available at* https://www.financialresearch.gov/public-appearances/2015/03/24/financial-regulation-summit/.

12. Several other SBR implementations are under way, not covered in detail by this paper. In Finland, Vero Skatt – the Finnish Tax Service – introduced in 2016 joint filings with the National Board of Patents and Registration. *See* "Tilinpäätös 2.0 ensimmäisten käyttöön syyskuussa 2016," June 21, 2016, *available at* https://www.vero.fi/fi-FI/Tietoa_Verohallinnosta/Uutiset/Tilinpaatos_20_ensimmaisten_kayttoon_syy(40033). Authorities in Sweden embarked in 2016 on a project coordinated by the Swedish Business Register – Bolagsverket – that aims to introduce Standard Business Reporting across several government organizations, including Statistics Sweden, the Swedish Financial Stability Authority, and others. See "Information days about digital filing of annual reports," Bolagsverket, November 28-29, 2016, *available at* http://www.bolagsverket.se/polopoly_fs/1.13824!/bolagsverkets-presentation.pdf. In 2015, South African regulators jointly developed a governance framework and nationwide standard to guide all relevant regulators in the consistent implementation of open data. The framework has already been applied in practice by the South African business register. *See* "XBRL South Africa Standard," Companies and Intellectual Property Commission, 2015, *available at* http://www.cipc.co.za/files/3414/3221/4863/BR-AG_XBRL_SA_Standard_Architecture_Document_-_PUBLIC_DRAFT_V1_01.pdf.

13. *See* Paris Cowan, "Tax office claims $1 billion in savings from SBR," *ITNews*, August 3, 2016, *available at* www.itnews.com.au/news/tax-office-claims-1-billion-in-savings-from-sbr-432460.

REFERENCES

14. Organisation for Economic Co-operation and Development, Forum on Tax Administration: Taxpayer Services Sub-Group, *Guidance Note: Standard Business Reporting*, July 2009, *available at* www.oecd.org/tax/forum-on-tax-administration/publications-and-products/admin/43384923.pdf ("OECD Paper"), at 9.

15. *See, e.g.*, Data Coalition, "Coalition to SEC: Replace Corporate Disclosure Documents with Open Data," July 22, 2016, www.datacoalition.org/coalition-to-sec-replace-corporate-disclosure-documents-with-open-data/ (describing SEC's persistent use of unstructured documents to collect disclosures from publicly-listed companies).

16. OECD Paper at 9.

17. Logius & Thauris, *Challenging the Chain*, March 2015, *available at* challengingthechain.com, at 381.

18. *Id.* at Appendix A.

19. *Id.*

20. *Id.*

21. *Id.* at XVIII.

22. Interview with Frans Hietbrink.

23. *Challenging the Chain* at Appendix A.

24. *Id.*

25. Interview with Frans Hietbrink.

26. "The use of SBR would only be mandatory for declarations that reporting parties (businesses or their intermediaries) sent in directly from software packages, thus system-to-system. For that reason, this step is referred to as 'mandating' the use of SBR. Alternatives, such as submitting reports via a portal on the Tax and Customs Administration website, will nevertheless continue to exist." *Challenging the Chain* at Appendix A.

27. *Id.*

28. *Id.*

29. Ministerie van Economische Zaken, January 2017, *available at* https://www.kvk.nl/download/QA_deponeren_jaarstukken_jan2017_tcm109-431972.pdf (in Dutch). *See also* XBRL International, "New Mandate in the Netherlands," November 29, 2015, www.xbrl.org/news/new-mandate-in-the-netherlands/.

30. *See* Data Coalition, "Dutch Government Envisions Business Reporting in Zero Clicks," December 10, 2015, www.datacoalition.org/dutch-government-envisions-business-reporting-in-zero-clicks/.

31. See XBRL International, "SBR Expands to Credit Reporting in the Netherlands," February 12, 2015, www.xbrl.org/news/sbr-expands-to-credit-reporting-in-the-netherlands/.

32. Fujitsu, *SBR Programmes: Reducing administrative burden through countrywide standardisation*, October 2016, *available at* www.fujitsu.com/global/documents/products/software/middleware/application-infrastructure/interstage/download/xbrl/FUJITSU-SBR-Programmes.pdf.

33. *Id.*

34. *Challenging the Chain*.

REFERENCES

35. *Id.* at 40.

36. "Digitaal aanleveren jaarcijfers MKB ook bij ING de norm," ING, October 2014, *available at* https://www.ing.nl/nieuws/nieuws_en_persberichten/2014/10/Digitaal_aanleveren_jaarcijfers_mkb_ook_bij_ing_de_norm.html.

37. Australian Government, *Rethinking Regulation: Report of the Task Force on Reducing Regulatory Burdens on Business: Australian Government's Response, 2006, available at* archive.treasury.gov.au/documents/1141/PDF/Reducing_Regulatory_Burdens_on_Business_Final_Government_Response.pdf, at i.

38. *Id.*

39. *Id.* at Recommendation 5.11.

40. Australian Government, *Australian Reporting Dictionary, available at* dictionary.sbr.gov.au/ (accessed February 8, 2017).

41. Australian Prudential Regulatory Authority, *Adopting Standard Business Reporting: A Guide to Adopting SBR for APRA Reporting*, August 2011, *available at* http://www.apra.gov.au/Super/Documents/AdoptingSBRforAPRA reporting.pdf, at 1.

42. *Id.*

43. Australian Government, *Standard Business Reporting Four Years On*, June 2014 ("*Four Years On*"). One key aspect of Australian SBR that goes beyond regulatory reporting is a functionality for private employers and retirement funds to use the SBR system to automatically check an employee's details against official tax details maintained by the Australian Tax Office. *Id.* at 3.

44. Id. at 9.

45. Australian Government, *SBR Board Meeting Item 4.0*, April 8, 2016, at 4.

46. *Four Years On* at 6.

47. Interview with Karen Lay-Brew.

48. Interview with John McAlister.

49. Australian Government, *Report of the Australian Business Registrar, 2014-15* ("ABR Report), *available at* abr.gov.au/uploadedFiles/Content/Download_Files/ABRreport2014-15.pdf, at 3.

50. Australian Taxation Office and Deloitte, *ABR Program Savings Review* ("Deloitte Report"), March 24, 2016, at 26.

51. *See, e.g.*, XBRL International, "Sweden Looks to SBR," June 3, 2016, www.xbrl.org/sweden-looks-to-sbr/.

52. The most prominent endorsement of U.S. SBR can be found in the Center for Data Innovation's 2016 presidential transition report. Center for Open Data Enterprise, *Open Data Transition Report: An Action Plan for the Next Administration*, October 2016, *available at* http://reports.opendataenterprise.org/transition-report.pdf, at recommendation 24.

53. Barry Burr, "Stanford Center's New Database," March 20, 2006, *Pensions & Investments*, www.pionline.com/article/20060320/PRINT/603200720/stanford-centerx2019s-new-database.

54. XBRL US, "DTCC, SWIFT and XBRL US Join Together on Plan to Improve Processing of Corporate Actions Announcements," May 28, 2009, www.prnewswire.com/news-releases/dtcc-swift-and-xbrl-us-join-together-on-plan-to-improve-processing-of-corporate-actions-announcements-62005092.html.

REFERENCES

55. *See* XBRL US, "Corporate Actions," xbrl.us/home/industries/corporate-actions/ (accessed February 8, 2017).

56. Emilia David, "FIBO Marches Forward: A Look Inside State Street's FIBO Proof of Concept," June 3, 2016, *Waters Technology*, www.waterstechnology.com/waters/feature/2459451/fibo-marches-forward-a-look-inside-state-streets-fibo-proof-of-concept.

57. 44 U.S.C. 35.

58. Public L. No. 111-203, Title I.

59. *See* NIEM.gov (accessed February 28, 2017).

60. Interview with John Truzzolino.

61. H.R. 2477 (114th Congress).

62. Pub. L. No. 113-111.

63. Interview with Danny Kermode.

64. Jessica Davis, "Big Data, Analytics Sales Will Reach $187 Billion By 2019," May 24, 2016, *Information Week*, www.informationweek.com/big-data/big-data-analytics/big-data-analytics-sales-will-reach-$187-billion-by-2019/d/d-id/1325631.

65. Kevin Roose, "Is Silicon Valley the Future of Finance?," June 1, 2014, *New York*, nymag.com/daily/intelligencer/2014/05/is-silicon-valley-the-future-of-finance.html.

ABOUT THE AUTHORS

Michael Middleton is a Certified Anti-Money Laundering Specialist (CAMS) and a former Chief Compliance Officer and works in financial services and enterprise product management. This paper is an independent collaboration with the Data Foundation, and the views and opinions expressed here do not necessarily reflect the views of his current or past employers, their affiliates, partners or employees.

Joseph Kull is a Director in PwC's Washington Federal Practice focusing on public sector budget, grants, and financial management and performance. He retired after 32 years of Federal service, including 20 years in the Senior Executive Service. His last four years were at the White House Office of Management and and Budget (OMB) as the Deputy Controller for federal financial management, responsible for accounting, auditing, financial systems, internal controls and grant management for all Federal agencies.

Prior to OMB, Kull spent over 15 years as the Chief Financial Officer and Budget Director at the National Science Foundation, and 11 years as Comptroller and Budget Director of the Civil Aeronautics Board. He began his career as an auditor with Arthur Andersen & Co.

Kull has an MBA and CPA, CFGM (Certified Government Financial Manager), and CGMA (Certified Global Management Accountant) credentials. He is a member of the American Institute of CPAs (AICPA), the Association of Government Accountants (AGA), and a number of professional and advisory boards and committees. He has testified before Congress, has taught accounting at the college and graduate level for 25 years, and was a member of the Financial Accounting Standards Advisory Board (FASAB).

Michal Piechocki is the co-founder, co-owner and CEO of the Business Reporting – Advisory Group, leading its global operations and advising international governmental and supervisory institutions on the adoption of international data standards, methodologies, architectures and technologies. His expertise contributed to projects in Europe, Australia, Asia and North and South America.

Michal serves as a Member of the Board of Directors of XBRL International ,overseeing development and worldwide adoption of the electronic XBRL standards, and a Member of the IFRS Taxonomy Consultative Group (IASB London), where he contributes to the strategy of the IFRS XBRL Taxonomy development and international implementation. Michal chairs the Frankfurt Group Technical Workshop on Optimisation of Data Standards hosted by the European Central Bank.

Prior to his work at BR-AG, Michal was a Member of the International Accounting Standards Committee (IASC) Foundation's XBRL Team, developing the International Financial Reporting Standards (IFRS) XBRL taxonomy and related projects. He graduated from the Poznan University of Economics, Poland, where he received a Master's degree in economics. Michal is an author of a number of publications, including a joint publication for IFAC and ISACA, and a book, XBRL for Interactive Data: Engineering the Information Value Chain, and a speaker at variety of financial, technological, supervisory and statistical conferences.

Hudson Hollister founded the Data Coalition in 2012 to serve as the leading industry advocacy group for open data in government. He has helped to craft landmark U.S. government data reforms, including the Digital Accountability and Transparency Act of 2014 (DATA Act) and the Financial Transparency Act. In 2016, he founded the Data Foundation to support open data transformations through research, education, and outreach. Hollister received the FCW Federal 100 award in 2015 and was recognized as a member of the Fastcase 50 and the LinkedIn Next Wave in 2016.

Prior to founding the Data Coalition and the Data Foundation, Hollister served as counsel to the Committee on Oversight and Government Reform of the U.S. House of Representatives and as an attorney fellow in the Office of Interactive Disclosure at the Securities and Exchange Commission. Before his government service, he was a securities litigator in the Chicago office of Latham & Watkins LLP.

Hollister received his Juris Doctor from Northwestern University School of Law in 2005 and a bachelor's degree in political science and mass communication from the University of St. Francis in 2002. He is a native of Joliet, Illinois.

Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

Creating a Linchpin for Financial Data: Toward a Universal Legal Entity Identifier

John A. Bottega and Linda F. Powell

2011-07

Creating a Linchpin for Financial Data: Toward a Universal Legal Entity Identifier

Discussion Paper: Approach toward establishing and maintaining a universal standardized legal entity identifier, or LEI, for financial data.

John Bottega, Chief Data Officer, Markets, Federal Reserve Bank of New York
Linda Powell, Chief, Micro Statistics Section, Division of Research and Statistics, Board of Governors of the Federal Reserve System

I. Introduction

The financial industry runs on information and data. Although financial data are made up of innumerable complex and idiosyncratic components, a fundamental building block for analysis is reference data about companies, organizations, and firms (henceforth referred to collectively as entities). Reference data might include a number of things, but an essential component is a systematic structure or code that uniquely identifies entities and their legal relationships with parent companies and subsidiaries capable of tracking changes in these relationships over time and quickly incorporating information on newly created entities. A universal, standard legal entity identifier (LEI) would likely provide a "public good" in that it could permit cheaper and more efficient analysis for all interested parties. It could also facilitate analysis that is currently incredibly difficult due to the plethora of proprietary entity identifiers.

An LEI could also be a critical component for measuring and monitoring systemic risk. The financial crisis demonstrated the extreme complexity of interrelationships and dependencies that exist between parties, counterparties, issuers, guarantees, and guarantors and how strains can rapidly spread through the financial network when one or more of the nodes within these horizontal or vertical relationships come under pressure. In principle, a system of unique identification of every entity would help to map these types of (inter)relationships in the financial system and allow a better understanding of the key linkages in advance of a crisis.

The financial services industry has been exploring the issue of unique entity identification for decades. More recently, several efforts have been made to advance the idea of a standard LEI, but competing priorities, funding issues, and an evident lack of industry focus have kept such a standard, and the benefits it could have yielded, from being implemented. However, recent economic events (historic market turmoil resulting in unprecedented numbers of mergers, acquisitions, divestitures, bankruptcies, and so on), suggest that this may be the right time to reconsider the usefulness of such a standard, especially one that is universal and based on an open architecture and determine how a universal and standardized LEI could be implemented. Indeed, the passage of the Dodd–Frank Wall Street Reform and Consumer Protection Act requires several financial regulators to write rules that involve entity identification, and the legislation set a tight time line for establishing these rules.

This paper explores the current state of entity identification in the market, the problems generated by the fragmentary nature of the current systems of identification, and the best-practice options for entity identification. To address these issues and move the discussion forward, the paper lays out a collaborative approach to the way in which the financial regulators and the financial industry might build and maintain a system of LEIs that we think provides incentives to attract interest for the wide range of parties involved. Standardizing identification of organizational and instrument structures identification are equally important, but this paper focuses on the best practices surrounding the implementation of an LEI at the entity level because organizational hierarchy, instrument identification, and counterparty exposure all build upon entity identification, therefore requiring that the LEI problem be solved first.[1]

[1] Organizational hierarchy refers to corporate structure, including ownership and affiliation. Determining the rules and requirements will be complex and will need to be vetted with financial industry professionals and the regulatory

II. Current State of Entity Identification

The ability of a financial institution to uniquely and precisely identify, define, and link business entities is critical to a wide array of essential business and risk-monitoring processes. For example, business functions such as sales (that is, a holistic view of the client), compliance (for example, "know your customer" requirements), and risk management all rely on unique entity identification. Regulators may require a similar degree of precision as they assess the financial health, systemic risk, and other aspects of markets and their participants as part of their statutory responsibilities and practices.

Although private, public, and vendor entity identifiers are in use today, there is no single or tightly integrated identifier that is consistent across all sectors. There is also no consistent representation of an entity's organizational structure that is commonly used or universally available. Many institutions and agencies cross-reference their identifiers to one another, but ambiguities and inconsistencies in those relationships often make cross-referencing difficult and inaccurate. Simply put, having a multitude of identifiers adds layers of complexity, increases the potential for errors, and results in redundant efforts.

Within the Private Sector

Within the private sector, entity identification touches so many aspects of companies' critical business functions that many firms have created their own internal identifiers to facilitate their business objectives. Even within the same firm, many of these internal solutions have been developed on a department-by-department or function-by-function basis, further complicating internal business flows. In the cases where internal solutions may have provided some relief, on an aggregated, industry-wide basis, these stop-gap measures have further aggravated and complicated an already disparate, inconsistent, and incompatible industry-wide entity identification infrastructure.[2]

Within the Public Sector

The public sector, especially financial and securities regulators, have had to develop identifiers over the decades to track the entities they supervise. However, the identification schemas are often incomplete, do not include all financial organizations, and include relatively few nonfinancial organizations.

The ID_RSSD is the primary identifier for the Federal Reserve's National Information Center (NIC). The NIC includes entity identifiers (the ID_RSSD) and codes for organizational hierarchies. For example, for each bank holding company maintained in NIC, details on all the entities in which the bank holding company has a regulatory or controlling interest are included. These entities include the holding companies themselves (the ultimate parent) and the banks, branches of banks, and nonbank subsidiaries associated with the ultimate parent. Having the ID_RSSD as a unique identifier, combined with the organizational hierarchy, is a powerful tool that allows regulators to sift through a complex web of corporate holdings. Although the ID_RSSD is used for regulatory reporting and is used by some data vendors, its

community. Similarly, instrument identification refers to the need to develop standards for tradable instruments that allow for tracking and auditing through an instrument's life.

[2] Reference Data User Group, Entities and Funds Committee (2003), "Legal Entity Identifiers," Discussion Paper, April 22, http://archive.fisd.net/referencedata/20030422rduglei.doc.

coverage is not complete beyond the banking industry and is not widely used outside of the regulatory community.

Other commonly used identifiers in the public sector include the CIK (Central Index Key), established by the Securities and Exchange Commission (SEC), which is used to identify issuers and certain shareholders in the SEC's EDGAR system; the Financial Industry Regulatory Authority's Web CRD (Central Registration Depository) and IARD (Investment Adviser Registration Depository) identifiers, which are used to identify broker–dealers, investment advisers, and investment adviser agents; and the National Association of Insurance Commissioners' company code and a group code which is used for the insurance industry.

When looking at this data landscape, the challenges for data analysis become obvious. When combining data collected from these separate industries, researchers and regulators must perform complex and time-consuming data matching across identifiers. Any researcher trying to merge banking data with data from other agencies must create and maintain his or her own cross-references between the data sets. And since, in many cases, the only common "link" between data sets is the entity name, matching on a name can be extraordinarily time consuming and can easily lead to erroneous results, particularly where no common, rigorous naming convention for entities exists.

In the Financial Data Vendor Industry

In the financial data vendor industry today, many vendors offer entity identification numbers and hierarchies as part of their product offerings. Many vendors adhere to industry best practices, providing unique identification of companies over time (see section III of this paper). However, most of these identifiers and hierarchies are proprietary and restricted contractually as to their use and redistribution. Many vendors see proprietary identifiers as a means to be commercially "sticky" with their clients creating a commercial conundrum that does not lend itself to what should be an industry-wide open standard solution.

Other Key Industry Players

Several other key players should be considered in evaluating the current state of all existing legal entity identification solutions. Several industry utilities—the Society for Worldwide Interbank Financial Telecommunication (SWIFT) or the Depository Trust and Clearing Corporation (DTCC), for example—may be willing to provide expanded entity identification solutions for the market. The International Standards Organization, or ISO, maintains various unique identification standards and is considering expanding existing standards or creating new ones to accommodate this industry need. And finally, several vendors have recently announced to the industry that they are willing to offer or are offering "open standard" identification solutions—the terms, conditions, and commercial aspects of these offerings have yet to be fully investigated.

Summary

In summary, the current landscape of legal entity identification solutions is disparate and incomplete and does not consistently meet the needs of the finance industry or its regulatory agencies. Understanding the current products and services available in the industry today is critical to understanding the operational and risk challenges that the industry faces, and it is also critical to formulating a best-practices solution that can provide both the private and public sectors with a way to foster improved efficiency and improved risk oversight.

III. Value of a Standardized LEI

As previously stated, unique legal entity identification is a critical factor of input to operational efficiency and risk management. Without an unambiguous and persistent identifier, the industry (and the regulatory community) faces operational hurdles on a regular basis. This section uses a number of specific examples to demonstrate that a case can be made that a universal, standardized LEI has the key aspects of a "public good"—that is, by allowing efficient analysis of firm-wide or industry-wide financial activity, a standardized LEI essentially offers positive externalities to the research and regulatory communities.

Below are examples of the hurdles encountered because of the lack of an industry-wide LEI:

1. *Identification of Non-Broker-Dealer Financial Industry Affiliates and Parent Companies*
   One of the key responsibilities of the regulatory community is to analyze the risk and effect of broker–dealer firms. In performing this analysis, a review of parent and affiliate companies of the broker–dealers is often required.

   Within the regulatory community, CRD numbers are used to identify broker–dealers. For broker–dealers, CRD numbers enable easy identification of parent and affiliates. However, if the parent, affiliates, or both are not broker–dealers, then outside data (such as the SEC form 10-K) must be used.

   Since no unique identifier is used across disparate sources (in this case, on the broker–dealer reports and the SEC financial 10-K and 10-Q reports), the parents and affiliates must be identified manually. Today the non-broker-dealer, non-investment-adviser affiliates and parent are identified by name, and as previously discussed in this paper, manual identification based on nonstandard naming conventions is highly prone to error.

   The introduction of an LEI could allow for consistent identification of the same non-broker-dealer, non-investment-adviser affiliate or parent associated with multiple broker–dealers and provide greater ability to identify systemic risk in broker–dealer firms.

2. *Identification of Counter Parties Involved in Trading, Clearing, and Settlement Activity*
   a. *Exchange product transactions*
      In today's market, each exchange (for example, the New York Stock Exchange, London Stock Exchange, and NASDAQ Stock Market) assigns a different market participant identifier to each broker–dealer. Often within the same exchange, local exchange rules allow broker–dealers to use multiple market identifiers for the same participant.

      The use of multiple market participant identifiers and the lack of consistent identifiers across exchanges combine to create a process that, in some instances, presents complications in the identification of the broker–dealer responsible for trades in a consolidated order audit trail.

      The introduction of an LEI would allow for the determination of the broker–dealer that facilitated the transaction and assist in assessing compliance with SEC and exchange rules.

   b. *Over-the-counter product transactions*

For over-the-counter (OTC) product transactions, most broker–dealer firms have developed internal solutions, sometimes across multiple systems within the same organization, to identify the counterparties to a transaction. Without a unique identifier assigned to each of the counterparties, it is extremely difficult to identify the exposures of the parties resulting from the transactions.

An LEI could also aid in position tracking. For each commodity, commitments of traders' reports issued by the Commodity Futures Trading Commission (CFTC) provide information on the size and direction of the positions taken, across all maturities, by three categories of futures traders: "commercials," "noncommercials," and "nonreportables."

The introduction of an LEI could enable easy identification of the same party across multiple transactions and, as a result, identify exposure and aid in position tracking activity.

c. *Instrument issuance tracking*

Instrument issuance tracking refers to the tracking of financial instruments (also known as issues) issued by legal entities (also known as issuers). In order to effectively maintain the relationships of "issues to issuers," and track these issues in the secondary market, unique identification of both issue and issuer are required.

Although this paper is focused on discussing the LEI (identifier of the issuers), it is worth noting that significant gaps still exist in the assignment of financial instrument identifiers. Although multiple financial instrument identifiers are in use today (CUSIP, ISIN, VALOREN[3], and so on), many asset classes are still without standard identifiers (for example, certain derivatives, loans, and so on). In order to establish a robust linkage connecting parent to child and child to issue, both instrument and legal entity standards should be established, made consistent, adopted, and supported.

This work is also critical to tracking changes due to corporate actions. Tracing entities (and their issues) through corporate actions can be difficult, especially with regard to small entities that are not covered by analysts or whose reports and announcements are not disaggregated by data intermediaries. The use of an LEI, traceable throughout the life of such an entity, could improve the ability of investors and regulators to track such activities.

d. *Payment, clearing, and settlement activity*

Once counterparties have entered into a financial transaction, they must clear and settle that transaction. This process often involves not only the original counterparties to the transaction but also a host of intermediaries and financial market utilities, including a variety of financial institutions. As with exchanges, each financial market utility assigns its own participant identifier(s) to each clearing entity.

The use of multiple clearing participant identifiers and the lack of consistent identifiers across payment systems, central securities depositories, and central

[3] CUSIP (Committee on Uniform Securities Identification Procedures), ISIN (International Securities Identification Number), and VALOREN numbers are codes that uniquely identify specific securities issues. CUSIP is used primarily in the U.S. and Canada. VALOREN is used primarily in Switzerland, and the ISIN is used on internationally traded securities.

counterparties present complications in the identification of common legal entities participating across multiple financial market utilities and their related settlement activity and exposures, both on an ongoing basis and in times of financial stress.

The introduction of an LEI would facilitate identification of the same party across multiple financial market utilities and, as a result, facilitate the identification of common entities and aid in tracking settlement activity and exposures.

3. *Economic Research*

Researchers at government agencies, at universities, and in the private sector frequently need to combine data from a variety of sources in order to get a comprehensive picture of a particular market or profile a particular entity. When combining data from multiple sources, it is critical that data collected be normalized, combined, and compared.

In the absence of a standard LEI, researchers, as well as financial institutions, are forced to perform time-consuming and costly cross-referencing, mapping, and reconciling exercises before they can effectively analyze the data collected, enabling them to provide the necessary oversight over complexity and guard against unacceptable risk. In the aftermath of the recent market crisis, greater focus will be placed on such analysis, some of which is mandated by new legislative law. Standardized LEIs are critical enablers to this type of analysis.

It is also necessary for researchers to look at entities or panels of entities over long periods. In constructing time series, the legal entity is not always the entity of greatest interest to researchers. For instance, it is not uncommon that a merger of banking organizations results in one legal entity acquiring a bank charter and negligible assets while another legal entity (operating under a different charter) acquires the bulk of the assets and liabilities. In some cases, researchers prefer to follow the entity holding the assets and liabilities rather than the entity that acquired the charter. In either analysis, clear and unambiguous identification of entities is the elemental building block that enables analysis and tracking of legal entities over historical periods.

In the end, macroprudential regulation and policy decisions are driven by economic research, so the ability to collect and properly analyze data from across the industry has a direct effect on these decisions. Improved economic research through the implementation and use of a standard LEI could lead to a more effective regulatory regime and better informed policy decisions.

4. *Holistic View of the Business*

The challenge of entity identification is not limited to regulators and researchers.[4] Private organizations and data vendors that consume, create, aggregate, or store data about financial firms also have a need for industry-wide entity identifiers. Many critical business functions within financial firms and financial market utilities are dependent on unique identification of legal entities. As firms have migrated away from the traditional product-centric operational strategies to a more service oriented approach, businesses are increasingly analyzing a diverse set of product offerings across business lines rather

[4] David Bannister (2010), "Single Customer View: Keeping One Eye on the Ball," *Banking Technology,* September 7, https://bankingtech.com/bankingtech/single-customer-view-keeping-one-eye-on-the-ball/20000186662.htm;jsessionid=343059201690E8054AC34A446423717C.f11b1cefac76ad95c7627468fee9bde7e866d022.

than evaluating products in isolation. This is referred to as a holistic view. The need for a holistic view is also true of financial market infrastructures such as central counterparties and settlement systems that are increasingly becoming more integrated through operational linkages and common corporate relationships.

In order to achieve the holistic view, firms are constructing consolidated views (building central data warehouses), where information from multiple business lines are pooled together for analysis and review. As stated previously, generating and maintaining internal identifiers is a costly and error-prone exercise. And the ability to share data across firms, or report to regulators in a consistent and standard manner, is hampered by the need for additional cross-referencing.

Having publicly available LEIs could enable organizations to operate more efficiently; could enable organizations to provide better risk analysis and customer service; and could better prepare organizations to be compliant with regulatory reporting requirements resulting from newly implemented regulatory reform.

IV. Key Elements of a Standardized LEI

When creating a universal, standardized LEI, it makes sense to follow the best practices that have been established in the development of proprietary identifiers. Some of the key components of what should be considered in defining the LEI standard are as follows:

1. *Scope of Coverage*

   All eligible market participants, including governmental agencies such as the Federal Deposit Insurance Corporation, or infrastructure participants such as the DTCC, should be assigned a unique LEI. These participants include, but are not limited to, financial intermediaries (banks and finance companies), companies listed on an exchange, companies that trade stock or debt, entities under the purview of a financial regulator, and their holding companies.

2. *Entity Types That Need Identifiers*

   Entity types should include issuing firms, entities acting as guarantors, selling firms (broker–dealers), buying firms (asset managers), clearing and settlement organizations, custodian and agent banks, payment system participants, distributors of financial products, exchanges and other trading system operators, collective investment vehicles and portfolios, hedge funds and fund managers, partnerships, government bodies , and supranational organizations.

   Although the need to identify subsections of a firm, such as a branch or trading desk, is sometimes necessary for use by regulators or market participants, the need is not uniform, and therefore the LEI should be set at the entity level. In the cases where a corporate hierarchy exists, the LEI should be assigned to each entity within the organization, not just the parent. Estimates indicate that within the United States, this universe would total between 500,000 and 2,000,000 entities.

3. *Structure of the Identifier*

   Several characteristics of the LEI are ideal for the identifier to be useful to a large audience.

   a. *Singularity and uniqueness*

There should be only one identifier per entity. Each entity within a corporate organization should have its own unique identifier. And every identifier should be unique and never reused.

Singularity and uniqueness are necessary to ensure that data users can confidently and easily identify a specific organization. Singularity would require that, over time, financial regulators would recognize the LEI even if they continue to maintain a separate internal identification system.

b. *Persistence and neutrality*
An identifier should follow an entity through its life regardless of corporate actions or other business or structural changes. The LEI should follow an entity through name changes, location moves, charter changes, and the acquisition of other entities. Persistence is important not just because it reduces the need to research changes but also because it reduces errors in analysis. Almost all economic analysis includes some evaluation of data over time.

For an identifier to be persistent over time, it should be neutral. For example, it is popular to incorporate geographic information or company name information into a corporate identifier. However, this practice violates basic data management principles and best practices regarding unique identification symbology. Descriptive attributes should not be coded into the identifier. Doing so creates a tight coupling of identifier to characteristics and, if allowed to happen, requires an identifier to change every time an entity characteristic changes. Entity characteristics should be viewed as separate elements within a reference data system and should not be incorporated into the identifier.

Only in the case where the legal status of an entity changes should the assignment of a new identifier be considered (usually in the case of a major corporate action, such as a merger or acquisition), and it should follow very strict and comprehensive rules. If this event occurs, the changing of the LEI should be based on a set of principles defined by a council of rule makers who are responsible for maintaining historical identifiers and links to those identifiers for audit and historical analysis purposes.

c. *Extensibility*
To ensure that the LEI will be persistent and unique over time, it is important that the LEI be extensible. The identifier should be robust enough to allow for growth in the volume of identifiers without having to reuse numbers. To ensure extensibility, standard algorithms used in industry today to create and properly size unique identifiers should be used in creating the LEI.

d. *Reliability and interoperability*
Finally, if the LEI is expected to be widely accepted, assurances that it is reliable and interoperable must be made. The mechanism for assigning and maintaining identifiers must ensure high quality. Users of the LEI must be confident that they have uniquely and accurately identified the firm they are looking for.

Where possible, the LEI should be compatible with existing systems and not conflict with other numbering or identification systems. The LEI must be usable in different computer environments to facilitate automating processes. In addition, the schema should be standard and work across various platforms.

4. *Public Availability*

   Use of identifiers must not be contractually restricted in their use. The LEI must be available for use for report collection and dissemination. It is particularly important that LEIs be publicly available for counterparty reporting and identification.

5. *Incentive Compatibility*

   To the highest degree possible, entities should desire to use the identifiers in their accounts, for payments, for risk management, and for other purposes and to act in ways that lead to maintenance of the system of identifiers. Incentive compatibility of the reference data system is critical if its usefulness is to survive inevitable shifts in market structure and function. If entities see an advantage in doing their part to maintain the system, the system is more likely to be robust than if it operates purely under compulsion. Some degree of compulsion may be necessary to start the process, but every effort should be made to involve players in such a way that everyone has a vested interest in its continuation.

6. *Registration Process*

   Time frames for assignment will need to be defined and the assignment of a new LEI should not materially hinder the normal course of a firm's business. The turnaround time for identifier assignment should be less than the average number of business days required to form a new organization and may need to be intraday in some cases.

7. *Quality Assurance Processes*

   Requirements should include a demonstration of high-quality processes—in identifier assignments and legal entity hierarchy mappings—and should demonstrate sound maintenance practices, especially throughout the corporate action event processes. These processes should be adequately governed and auditable.

   A critical quality control is ensuring that duplicate identification numbers are not erroneously assigned. The quality assurance processes should include checks for existing entities, including name searches, address searches, and combinations of text strings and other characteristics.

   A number of standards used in industry today ensure quality and accuracy in identification assignment. For example, the Item Unique Identification Standard is an identification assignment implemented by the Department of Defense (DoD) to uniquely and unambiguously identify objects (that is, equipment, operating materials, and supplies), enabling lifecycle traceability.[5] The Universally Unique Identifier is another unique identification standard used in software development, intended to enable unique identification without significant central coordination. No matter the methodology selected, LEI creation should adhere to industry best practices in identification assignment to ensure high quality and accuracy.

8. *Relationship to an Open Standard*

[5] U.S. Department of Defense, Defense Procurement and Acquisition Policy, "Unique Identification," webpage, www.acq.osd.mil/dpap/pdi/uid/index.html.

Entity identification should work through an open standard. As early as 1998, the U.S. government recognized the need to move away from institution- or government-unique standards toward voluntary consensus standards[6].

9. *Reference Data*

   Reference data should be sufficient to verify that users have correctly identified an entity. At a minimum, the reference data should include the entity's name and location and be part of the publicly available information.

V. Implementation Issues for a Standardized LEI

In addition to gaining consensus about what elements a standardized LEI should include, there would no doubt be numerous implementation issues to overcome. This section discusses three possible approaches to establish, implement, and service an industry standard LEI. The first approach suggests implementation via the private sector. The second suggests implementation by the public sector. The third discusses a hybrid approach, combining regulatory participation with industry infrastructure and best practices. For all three approaches, developing an industry-wide consensus on the approach will be difficult, and start-up costs are likely to be significant. A significant difference between the approaches is how decisions are made and who will bear the costs.

It should be noted that, regardless of the solution selected, a successful legal entity implementation and maintenance strategy will almost certainly involve some level of international cooperation.

1. *Private-Sector Solution*

   As previously stated in this paper, many vendors offer entity identification numbers and hierarchies as part of their product offerings. A number of vendors and industry utilities issue entity identification numbers today, including but not limited to Standard & Poor's, Avox, Omgeo, FactSet, Bloomberg, Thomson Reuters, Dunn & Bradstreet, Telekurs, Markit (red code), SWIFT, and Alacra. And many such numbers have been used successfully to facilitate trade settlement and cash transfer for years.

   For the private sector to be successful in establishing and implementing a unique LEI, a single identifier would need to emerge out of the multitude of identifiers that already exist. Although the industry has created a cache of artifacts, analysis, and design documents that speak to the problem, the obvious commercial and competitive challenges remain, as many vendors in the market view their identifiers as proprietary or as components of their larger product offerings.

   For the private sector to solve this problem itself, financial institutions would need to apply pressure to the vendor community and demand that a collaborative solution be reached.

[6] Office of Management and Budget (1998), *Circular A-119, Revised,* a circular on federal participation in the development and use of voluntary consensus standards and in conformity assessment activities, February 10, available at www.whitehouse.gov/omb/circulars_a119.

2. *Public-Sector Solution*

For the public sector to address this problem, the global regulatory community would need to become the assigning and maintaining agent of the LEI, operating across multiple jurisdictions. As implied in the private-sector approach, creating and maintaining a high-quality LEI system could be costly. For this approach to work in the public sector, government funding could be needed in every jurisdiction. The scope and size of the agency, given the responsibility of the LEI, could need to expand to accommodate the universe of entities that fall under the LEI banner and to support the community of users that need to acquire identifiers from the agency. And the agency should be prepared to interact on a regular basis with an expanded universe of global LEI facilitators.

The LEI could be viewed as and considered a critical public good. The most applicable examples of a public-sector solution, all-be them domestic, are the Social Security Administration's issuance of the Social Security number[7] and the Internal Revenue Service's issuance of the taxpayer identification number.

3. *Private-Sector Solution with Public-Sector Involvement*

The third approach is to establish a collaborative solution between the regulatory community and commercial providers.

In this model, the public sector acts as the guide and catalyst for the solution, while the private sector provides its expertise and vast data management infrastructure and distribution capabilities to establish and propagate the identifiers in the most efficient and effective manner possible.

Several existing models for public and private cooperation may be considered.

a. The first example of a successful implementation relates to Internet domain names. In June 1998, the Department of Commerce (DOC) issued a white paper endorsing the creation of a new not-for-profit corporation of private-sector Internet stakeholders to administer policy for the Internet name and address system. In November 1998, the DOC formally approved a new corporation, called the Internet Corporation for Assigned Names and Numbers (ICANN). Later that year, the DOC and ICANN established a memorandum of understanding (MOU). That MOU initiated a process intended to transition technical Domain Name System (DNS) coordination and management functions to a private-sector not-for-profit entity (that is, ICANN). The DOC retained a role with respect to the DNS via three contractual agreements.[8]

b. The National Information Exchange Model (NIEM) is another example of public and private cooperation. NIEM is designed to support processes and standards that allow jurisdictions to promptly and effectively share critical information. Data are formatted in a consistent manner so that they are exchanged and

[7] This example does not imply that the LEI would apply to all individuals.

[8] See U.S. Department of Commerce, National Telecommunications and Information Administration (1998), "Management of Internet Names and Addresses," statement of policy (Docket No. 980212036-8146-02), www.ntia.doc.gov/ntiahome/domainname/6_5_98dns.htm. Details regarding those agreements and Internet domain naming more broadly can be found, for example, in Lennard G. Kruger (2009), *Internet Domain Names: Background and Policy Issues* (Washington: Congressional Research Service), www.fas.org/sgp/crs/misc/97-868.pdf.

understood from organization to organization, without confusion derived from semantics. This standard allows NIEM to greatly assist in the coordination of efforts following large-scale emergencies by providing a shared information platform.[9]

c. A government-owned and contractor-operated (GOCO) model could also be considered. This model appears to be commonly employed by the military. The DoD contracts with hundreds of different companies that provide essential services for it at DoD-owned plants. Lockheed Martin and Boeing are examples of primary operators of GOCO plants, each with outstanding contracts of around $10 billion from the government in fiscal year 2009. This setup allows each party to more efficiently provide services for which they are well suited. The DoD outlines product needs, while contractors implement production steps.[10]

d. The automated clearinghouse (ACH) network is another example of private and public collaboration to meet the needs of the economy. Rules for the ACH network are set by the National Automated Clearing House Association. There are only two operators of the ACH system—the Electronic Payments Network, owned by The Clearing House, and FedACH, owned by the Federal Reserve System.

VI. Discussion

Recognition is spreading in the private and public sectors, both here and abroad, that standardized legal entity identification could serve as a critical element in the analysis and monitoring of financial stability and systemic risk.[11]

Upon reviewing the current state of legal entity identification in the industry today, looking at the gaps that exist and the challenges they create, and reviewing the possible approaches to addressing these challenges it appears likely that a private-sector solution with public-sector involvement may provide the most robust and expedient solution to this industry-wide problem. In addition to the practical advantage of a joint effort, this approach is also consistent with the practices defined by the OMB *Circular A-119, Revised,* which encourages public and private collaboration.[12]

---

[9] See National Information Exchange Model, "Learn More about NIEM," webpage, www.niem.gov/whatIsNiem.php.
[10] See FedSpending.org, a project of OMB Watch, "Contracts for Operation of Government-Owned Facilities—Government-Owned Contractor-Operated (GOCO) R&D Facilities (FY 2000–2009), webpage, www.fedspending.org/fpds/fpds.php?psc_sub=M181&detail=-1.
[11] Dick Hales (2004), "Who needs (or even wants) the new Legal Entity Identifier (LEI)?" IT-Director.com, June 17, http://www.it-director.com/business/content.php?cid=7166.

Office of Financial Research (2010), "OFR Policy Statement on Legal Entity Identifiers", http://www.treas.gov/ofr/docs/OFR-LEI_Policy_Statement-FINAL.PDF.
[12] The National Technology Transfer and Advancement Act of 1995 codified OMB *Circular A-119* and directs federal agencies to use voluntary consensus standards in lieu of government-unique standards except where inconsistent with law or otherwise impractical. This provision's intent is to eliminate the cost to the government of developing its own standards, decrease the burden of complying with agency regulation, provide incentives and opportunities to establish standards that serve national needs, encourage long-term growth for U.S. enterprises, promote efficiency

With the passage of the Dodd–Frank Act, various elements of this new legislation call for the implementation of data and content standards in collecting and analyzing critical market information. Recently, the CFTC and SEC released proposed rules addressing counter party identifiers.[13] Given this mandate, a second recommendation of this paper is to align the objectives of the establishment of a standard LEI with the data standard mandates of the act, to ensure consistency in approach, and to leverage the importance and urgency of these efforts to address these critical data needs.

The next steps toward resolving the LEI problem are to gather financial industry participants to explore the variety of issues. This work could include encouraging market participants to host information gathering sessions or rely upon regulators to develop a public process for examining these issues.

The approach should be open and collaborative. One method is to rely upon a series of requests for information by the involved regulatory community members that could gather input from interested parties on the various questions about structure, maintenance, governance, licensing, and the full range of issues.

The initial implementation should address the highest-priority use cases, with a road map to address the other use cases over time. The important topics of reference data and organizational hierarchy need to be addressed in conjunction with or shortly after the development of the LEI. To ensure robust reference and hierarchy data it could be beneficial if the implementation were flexible and iterative.

The plans to create and adopt a standard LEI must pass the rigors of industry acceptance and be viable and reasonable in its implementation. Throughout all of this and future analysis, iterative implementation should be considered wherever possible. Although this point is more of an implementation consideration than a specification consideration, moving this standard forward in a meaningful way through iterative rollouts, implementation, and acceptance could provide benefits more quickly and enable useful modifications to the standard.

---

and economic competition through the harmonization of standards, and further the policy of reliance upon the private sector to supply government needs for goods and services.

In addition, to promote trade and implement the provisions of international treaty agreements, the provision requires federal agencies to consider international standards in procurement and regulatory applications. As defined in OMB *Circular A-119,* "voluntary consensus standards" are standards developed or adopted by voluntary consensus standards bodies, both domestic and international. These standards include provisions requiring that owners of relevant intellectual property agree to make that intellectual property available on a nondiscriminatory, royalty-free, or reasonable royalty basis to all interested parties. "Voluntary consensus standards bodies" are domestic or international organizations that plan, develop, establish, or coordinate voluntary consensus standards using agreed-upon procedures.

[13] CFTC (2010), "Swap Data Recordkeeping and Reporting Requirements", http://www.cftc.gov/ucm/groups/public/@otherif/documents/ifdocs/federalregister112210.pdf.

SEC (2010), SEC Proposed Rules on Security-Based Swap Reporting", http://www.sec.gov/news/press/2010/2010-230.htm.

VII. Conclusion

Public and private industry has recognized for years that clear and unambiguous identification of legal entities is critical to financial research, markets monitoring, and systemic risk analysis. But the common problem encountered by all organizations that use financial data is that unique and accurate identification of legal entities and their subsidiaries without a recognized industry standard has been very difficult, costly, and prone to error.

The global financial community experienced what may have been the worst economic crisis since the Great Depression. And through that experience, many individuals now recognize the need for improvements in our global regulatory mechanism that will provide for real-time analysis across multiple financial markets to identify systemic risks and stresses in market conditions before they occur. For years, efforts to develop and implement an industry-wide legal entity identification standard have been unsuccessful. The economic incentive to invest in an operational standard was a difficult case for the industry to make. The vendor community tried to provide solutions for these private and public challenges; however, no solution has been sufficiently robust, comprehensive, or open to serve as an industry-wide standard. Viewing the LEI as a public good that could provide efficiencies across the financial industry and may help to create the incentives to develop an industry-wide standard.

A standardized and universal LEI could enable examiners, economists, and financial analysts to accomplish analyses during stressed market conditions and improve systemic analysis across the breadth of the financial markets. Such an identifier could improve analysis conducted not only by the regulatory community but also the financial services industry at large, both domestically and internationally.

Acknowledgments

The authors wish to thank their colleagues Thomas Allard, Nancy Brucks, and Robert Letson from the Board of Governors of the Federal Reserve System; Irina Leonova and David Taylor from the Commodity Futures Trading Commission; Alan Deaton and Mark Montoya from the Federal Deposit Insurance Corporation; Marty Colburn, Peter Oldershaw, and Elena Shuvalov from the Financial Industry Regulatory Authority; David Blaszkowsky and Matthew Reed from the Securities and Exchange Commission; Lewis Alexander, Michael Donnelly, and Adam Lavier from the U.S. Department of the Treasury for their research and contributions to this paper.

The views expressed in this paper are those of the authors and do not necessarily represent those of their organizations or the positions of fellow staff members within those organizations.

References

Bannister, David (2010). "Single Customer View: Keeping One Eye on the Ball," *Banking Technology* (September 7), "https://bankingtech.com/bankingtech/single-customer-view-keeping-one-eye-on-the-ball/20000186662.htm;jsessionid=343059201690E8054AC34A446423717C.f11b1cefac76ad95c7627468fee9bde7e866d022".

Data Management International (2010). *The DAMA Guide to the Data Management Body of Knowledge*. New Jersey: Technics Publications.

Hales, Dick (2004). "Who Needs (or Even Wants) the New Legal Entity Identifier (LEI)?" *IT-Director.com* (June), http://www.it-director.com/business/content.php?cid=7166.

International Swaps and Derivatives Association. (2010). Collateralized *Portfolio Reconciliation Best Operational Practices,* www.isda.org.

Office of Management and Budget (1998). *Circular A-119, Revised,* a circular on federal participation in the development and use of voluntary consensus standards and in conformity assessment activities, February 10, www.whitehouse.gov/omb/circulars_a119.

Reference Data User Group, Entities and Funds Committee (2003). "Legal Entity Identifiers," Discussion Paper, April 22, http://archive.fisd.net/referencedata/20030422rduglei.doc.

Tarullo, Daniel K. (2010). "Equipping Financial Regulators with the Tools Necessary to Monitor Systemic Risk," statement before the Subcommittee on Security and International Trade and Finance, Committee on Banking, Housing, and Urban Affairs, U.S. Senate, February 12, www.federalreserve.gov/newsevents/testimony/tarullo20100212a.htm.