Data Transparency Coalition
Comments on Establishment of Financial Data Standards (Data Exchange)
November 25, 2014

Submission to
Bureau of the Fiscal Service
U.S. Department of the Treasury
In Response to Request for Comments
Docket Number FISCAL-2014-0004

Table of Contents

I. Introduction

The Data Transparency Coalition (www.datacoalition.org) submits this paper to the Bureau of the Fiscal Service of U.S. Department of the Treasury (Treasury) in response to Request for Comments Docket Number FISCAL-2014-0004.

This section provides a brief introduction to the Data Transparency Coalition, its role in supporting the Digital Accountability and Transparency Act (DATA Act), and selected issues related to the DATA Act's implementation. Section II identifies the criteria that we believe will prove essential to success for the government-wide data standards that the DATA Act mandates. Section III proceeds to identify the key reporting regimes and concepts that make up the current landscape of federal spending reporting – providing a potential framework for DATA Act data standards. Section IV summarizes instructive examples of previous government data standards projects and initiatives. Section V directly addresses the questions contained in Treasury's Request for Comments. The responses in Section V include recommendations and lessons learned. Section VI provides a conclusion and summary.

The Data Transparency Coalition

The Data Transparency Coalition is the nation's only open data trade association.

Founded in 2012, the Coalition represents leading technology and consulting firms, including both industry leaders and growing startups. Some of our members offer software solutions and platforms. Others offer solution-agnostic expert advice.

The Coalition supports the publication of government information as standardized, machine-readable data, also known as open data.

We believe that the transformation of government information from disconnected documents into open data will facilitate public accountability, enable data-driven government management, and automate compliance. The transformation will also enable many of our members to pursue new business models and create high-tech jobs.

We believe that the transformation from disconnected documents to open data requires domain-specific policy mandates for data standardization and data publication. Through our advocacy, education, and collaboration, we seek to persuade government authorities to: (1) within each domain, adopt comprehensive and effective data standards; and (2) within each domain, consistently publish all information that is legally public.

To date, the Coalition has focused on supporting the open data transformation in two key domains of the U.S. federal government's information portfolio: spending and financial regulatory reporting. The DATA Act is a comprehensive mandate for open data in federal

spending. We hope to pursue comparable legislation to drive a similar transformation in financial regulatory reporting.

The Data Transparency Coalition's Role in the Enactment of the DATA Act

The DATA Act's enactment and successful implementation has been a primary motivator for the Coalition's founding and growth.

In 2011, Coalition founder and executive director Hudson Hollister drafted the original version of the DATA Act as counsel to the House Oversight Committee majority, serving committee chairman Rep. Darrell Issa. Though Rep. Issa and Sen. Mark Warner introduced the DATA Act in June 2011, they faced difficulty in convincing others in Congress and the executive branch of the need for a legislative mandate to standardize and publish federal spending information.

With the assistance of Runyan Public Affairs and with the early support of Teradata, RR Donnelley, and Level One Technologies, Hollister founded the Data Transparency Coalition in February 2012 to rally technology and consulting firms to support the DATA Act and similar proposals. The Coalition lobbied Congress for the DATA Act's passage and educated key constituencies on how data standardization and publication would benefit U.S. government and society. The Coalition also brought together industry leaders to demonstrate how software technologies could use federal spending data, once standardized, to deliver better accountability to citizens, better management to civil servants, and automated compliance to grantees and contractors.

The Coalition became the primary private-sector advocate for the DATA Act's enactment and successful implementation. In 2012 and 2013, our DATA Act Demo Days generated needed media attention. In 2013, Treasury announced its Intelligent Data Project, a precursor to its current implementation effort, at our annual conference. On April 28, 2014 – one day after the bill's final passage – our DATA Act Summit brought together the bill's sponsors, presumptive implementers, and the constituencies seeking to make productive use of federal spending data, once standardized.

The Importance of Treasury and OMB's Establishment of Data Standards

The DATA Act has the potential to transform the U.S. federal government and its relationship with the citizens it serves – but only if its crucial first step of establishing government-wide data standards for federal spending is successful.

1. Better accountability. First, if fully standardized and published, federal spending data will serve as a public resource for accountability. Spending data, once standardized, will be available for media, watchdog groups, and citizens to mine for valuable and unprecedented insights.

2. Better management. Second, if fully standardized and published, federal spending data will support new tools that allow government leaders to understand, analyze, monitor and predict. Such tools, sometimes described using the umbrella term *big data*,

improve the quality of decisions, and are in common use throughout the private sector already.

3. Automated compliance. Third, government-wide data standards for federal spending will support the automation of currently manual compliance tasks for recipients of contracts, grants, and other federal assistance. Automated compliance holds particular promise for the state and local government entities that receive the majority of federal grant funds.

We believe that the success of the DATA Act should be judged by whether accountability, management, and compliance improve. If the potential use cases in each of these areas that we discuss in Section V, below, are realized, the DATA Act should be considered a success. If not, then the DATA Act should not be considered a success.

All of these benefits are dependent on the government-wide data standards for federal spending whose establishment is required by the DATA Act. Under the DATA Act, Treasury and OMB must jointly establish these data standards by May 9, 2015, and are responsible for their maintenance thereafter. The data standards are the critical infrastructure for accountability, management, and automation.

We believe that Treasury and OMB are capable of establishing effective data standards for federal spending. In Section II, we discuss the criteria we believe will be essential to their effectiveness. In Section IV, we discuss examples of previous data standardization that may help guide Treasury and OMB in their task.

The Lack of Appropriated Funds for DATA Act Implementation

Significantly, the DATA Act was passed without the appropriation of additional funds to support its implementation by agencies (though the law does offset expenses by speeding Treasury's collection on delinquent loans).

We believe that the DATA Act, if properly implemented, will prove to be not an unfunded mandate but a fund-saving mandate for agencies and for the grantees, contractors, and aid recipients who must report to government. In theory, the DATA Act does not require a program budget, because it addresses business processes already performed in government. The law promotes the modernization and simplification of these processes because government-wide data standards for federal spending will enable the more efficient generation, collection, and exchange of such data. These changes will reduce the manual steps in disclosing spending to the public, conducting internal management, and reporting to government.

Nonetheless, the fund-saving benefits to agencies will not happen immediately upon enactment of legislation or the establishment of data standards. Although some of the implementation efforts discussed in Section IV, below, show that such benefits are realized very quickly, the DATA Act implementation effort will not be costless for agencies or recipients.

In the absence of an additional appropriation, therefore, the Coalition recommends that Treasury and OMB assist agencies to realign resources in strategic planning, human resources, information technology, procurement, grants management, and in programs across all the domains affected.

For example, we recommend that the implementation of the standards and the collection of standardized data be addressed in agency strategic plans. Another example is that participation in data standards governance activities by federal employees should be blended into existing position descriptions and performance reviews.

II. Essential Criteria for DATA Act Standards

The DATA Act requires Treasury and OMB to "establish Government-wide financial data standards for any Federal funds made available to or expended by Federal agencies and entities receiving Federal funds" (Sec. 4).

The technology industry uses the term "data standards" to refer to a wide variety of conceptual and technological structures. However, the DATA Act specifies just two types of standards: "common data elements" for federal spending concepts and a "widely accepted, nonproprietary, searchable, platform-independent computer-readable format" for the exchange of federal spending data.

The DATA Act's mandate for data standards covers all reports related to spending by the U.S. federal government, including both information reported by federal agencies and information reported by the recipients of federal funds. These reporting regimes are summarized in Section III.

Considering its wide reach, the DATA Act may represent the most extensive government data standardization effort ever attempted. However, the DATA Act is not unprecedented. The members of the Data Transparency Coalition have extensive experience with the implementation of data standards for government reporting – including regulatory, financial, and other forms of reports – throughout the world. Some of these projects have delivered their promised benefits; others have not.

In order to deliver the promised benefits of the DATA Act – better accountability, better management, and automated compliance – we believe the data standards must meet certain essential criteria.

The data standards themselves must be *complete*, *accepted*, and *nonproprietary*. The implementation of the standards must be fully *enforced, incremental*, and *sustainable*. The data exchanged through the system must be *supportive of validation*.

Essential Criteria for the Data Standards Themselves

1. Complete. To fulfill both the letter and spirit of the DATA Act, the data standards must comprehensively cover all information generated, collected, or exchanged by the

federal government related to its spending. The standards must not be limited merely to the subset of federal spending information that must be made publicly available on the USASpending.gov website. They must be capable of being used for all the spending-related reporting regimes affecting federal agencies, contractors, grantees, and other awardees. These reporting regimes are summarized in Section III.

"Complete" does not mean inflexible. Treasury and OMB need not establish a data element for every last concept generated, collected, or exchanged – only common concepts used by multiple agencies. Treasury and OMB should ensure that the data format they establish is easily extensible. (Extensibility is discussed in Section V.) The data standards should be designed for adjustment over time.

2. Accepted. The standards must be built on existing ones already accepted by government and industry. Data elements already available to identify common concepts, such as the Legal Entity Identifier (LEI) for entities, should be favored. Data formats already well-developed for government financial reporting, such as the eXtensible Business Reporting Language (XBRL), should be deployed where possible. Existing open-source work product, such as the Open Contracting Data Standard, should be reused. Standards already embraced by major stakeholders, particularly federal agencies and recipients, should be prioritized.

3. Nonproprietary. Federal spending data is public, so the standards governing such data must also be freely available. Treasury and OMB must reject data elements that are encumbered by licensing requirements. (The law specifically requires the data format to be nonproprietary, and includes no such requirement for the data elements, but it should not be read to encourage Treasury and OMB to select proprietary data elements.)

Essential Criteria for the Data Standards' Implementation

4. Incremental. Different agencies, domains, and reporting regimes should begin using the data standards incrementally. Since Treasury and OMB already collect agencies' financial, payment, budget, and grant reports, they may decide to apply the data standards to these reporting regimes first. On the other hand, Treasury and OMB are less directly engaged with agencies' contract reports – which are mostly managed by the General Services Administration – so they may choose to apply the data standards to contract reports only after successful adoption in other areas.

This does not preclude early participation in the development of the standards or observation as they are deployed. On the contrary, all major stakeholders should be involved as early possible in the implementation process.

5. Enforced. Once an agency, domain, or report has been made subject to the data standards, the standards must be fully enforced. As demonstrated by such ineffective implementations as the adoption of XBRL by the Securities and Exchange Commission, discussed below, deviation from data standards, uncorrected, undermines the quality and

usefulness of the data. Inadequate enforcement prevents the benefits of open data from being realized.

Therefore, for the reporting regimes they directly control, Treasury and OMB should ensure that reports out of compliance with the standards are rejected. For the reporting regimes they do not administer, Treasury and OMB should encourage collecting agencies to use their existing tools to enforce compliance. Consequences for errors in compliance must be predicable and proportionate.

6. Sustainable. The standards must be maintained for the long term. Treasury and OMB should adopt and establish a standards governance framework for that purpose. (We propose a governance framework below.)

Essential Criterion for the Data Exchanged Using the Standards

7. Supportive of Validation. The quality of federal spending data and veracity of any analysis will be greatly increased if submissions using the data standards can be easily validated against business rules. For every federal spending reporting regime, the responsible entity should create and enforce business rules that specify acceptable and unacceptable values. Treasury and OMB should establish standards with a view toward making business rule setting as quick and easy as possible. As discussed in Section V, Treasury and OMB should seek to strike a balance between document-based and rules-based validation.

III. Existing Reporting Regimes and Core Spending Concepts

As they approach the task of establishing data standards for federal spending, Treasury and OMB are not operating in a vacuum.

Through several key reporting regimes, the U.S. federal government already generates, collects, and exchanges extensive information on its use of taxpayers' and other funds. Many of these reporting regimes involve similar information, allowing us to discern several groups of core spending concepts that are used by multiple regimes.

The challenge facing Treasury and OMB – and the reason why the DATA Act was necessary in the first place – is that the existing reporting regimes are not coordinated with one another and the core concepts are expressed differently.

The DATA Act confers the necessary authority to address this challenge. For the first time, a single project can confront the federal government's whole structure of disparate spending-related reporting regimes. Government-wide data standards can bring together information from disparate reporting regimes without requiring the regimes themselves to be eliminated or consolidated, and without requiring existing systems to be redesigned. Government-wide data standards can align core concepts while still allowing regime-specific expression.

In order to be complete, DATA Act data standards must accommodate all of the existing reporting regimes and consistently reflect the core spending concepts.

Key Reporting Regimes

The DATA Act treats information reported by agencies and information reported by recipients (contractors, grantees, and recipients of other assistance) differently.

Agencies must conform the information that they report to the data standards within two years after those standards are established by OMB and Treasury. Recipient reporting will not immediately be made subject to the standards. Instead, the DATA Act requires OMB to conduct a pilot program to test the application of the standards to recipient reporting – and decide, after the conclusion of the pilot program, whether to require all recipients to conform the information they report to the standards.

Agency reporting. Agencies report their spending through five key reporting regimes: (1) Financial Account Balances, reported to Treasury; (2) Payment Requests, submitted to Treasury; (3) Budget Actions, reported to OMB; (4) Contracts, reported to the GSA; and (5) Grants and Other Assistance, formerly reported to the Commerce Department but now reported to Treasury.

Even within the same agency, the staff responsible for these reports often work in different places, have different professional backgrounds, serve different customers, use different systems, and apply different business rules. In a historical context, the separate regimes developed independently to meet different needs of the government.

Recipient reporting. Recipients primarily report their receipt and use of federal funds to the agency from which they received the funds. In addition, contractors, and some recipients of grants and other assistance, are subject to several registration and reporting requirements administered by the GSA. Finally, prime recipients – those who issue sub-grants or sub-contracts to sub-grantees or sub-contractors – must report their sub-awards to the FFATA Subaward Reporting System (FSRS), administered by the GSA on behalf of OMB.

Core Spending Concepts

All or most of the key reporting regimes reflect overlapping concepts, though they typically express them differently. Treasury and OMB have the opportunity to align different regimes' treatment of the following groups of core spending concepts: (1) transaction subject; (2) dollar amount; (3) payor/payee; (4) time; and (5) geospatial.

These groups have been recognized by the Department of Health and Human Services as the most important categories of data fields in current use for the tracking of federal spending.

(1) **Transaction subject** identifies the type of transaction, such as contract, grant or other award, or internal transaction. The transaction subject might also be a

budget line item or appropriated funds. The transaction subject answers the question: *what is being transacted and/or what is money being spent on*?

(2) **Dollar value** identifies the dollar values associated with a transaction, answering the question: *how much is being transferred or spent?*

(3) **Payor/Payee** identifies entities involved in a transaction, which could be an agency initiating or a recipient receiving the transaction. The payor/payee data answers the question: *who is transferring money to **whom**?*

(4) **Time** refers to temporal information, such as date of payment, period of performance, or accounting / budgeting period, answering such questions as: *when did payment occur? **When** did the work occur for which payment is made? To which accounting or budgeting periods does the transaction relate?*

(5) **Geospatial** refers to geographic and spatial characteristics such as place of performance. Geospatial information answers questions such as: *where was this the work performed and where were the payor and payee located?*

The development of data standards will require Treasury and OMB to choose consistent ways of expressing concepts from each of these groups across disparate reporting regimes.

For example, Treasury's collection of payment requests and GSA's collection of contract summaries both require agencies to identify the recipient of contract payments. Today the two reporting regimes express this information differently. Most payment requests submitted to Treasury identify the recipient using an Employer Identification Number (EIN). In contrast, contract summaries identify the contractor using a proprietary Data Universal Numbering System (DUNS) number. Through data standards, Treasury and OMB have the opportunity to harmonize these two reporting regimes so that payments from Treasury's regime can be matched against contracts from GSA's.

Another example is that agencies often purchase through other agencies' contract vehicles. A significant portion of federal procurement is conducted though indefinite deliverable / indefinite quantity (IDIQ) contracts in the Federal Supply Schedules program or through government-wide acquisition contracts (GWACs). Such contract vehicles make accountability difficult. But by aligning the ways in which core concepts are expressed, data standards can allow agencies and the public to track spending even when it is conducted through layers of contract vehicles.

Building Data Standards on Existing Regimes and Core Concepts

Treasury and OMB should build DATA Act data standards on the existing reporting regimes and fashion them to consistently reflect the core concepts.

We believe that the five groups of core concepts could become a government-wide core taxonomy, or basic data element dictionary, for federal spending. The creation of the

taxonomy would require input from those responsible for, and subject to, each existing reporting regime.

Therefore, we believe that Treasury and OMB should create an advisory body representing each existing reporting regime. These regime-specific advisory bodies could supplement the work of the government-wide DATA Act inter-agency advisory committee that Treasury and OMB have already established.

Working together, the inter-agency and regime-specific advisory groups could recommend first (1) a government-wide core taxonomy and then (2) regime-specific supplements to that core, each mapping to the core taxonomy. After the announcement of the first version of the standards by May 9, 2015, the regime-specific groups could be put in charge of harmonizing their own reporting regime with the core and with its own regime-specific taxonomy.

The example of purchases made through another agency's contract vehicles raised above may be instructive. This business reality many necessitate developing the standard and taxonomy to address use of various contract vehicles under the payors/payee group of elements. This would be a matter for the Contracts advisory group.

IV. Existing Data Standards Projects and Initiatives: Summary of Relevant Examples

This section summarizes short examples of relevant data standards projects and initiatives. We use these examples to address Treasury's specific questions more directly in Section V.

As the only trade association focused on open data, the Data Transparency Coalition has access to all its members' experiences in deploying and using data standards, only a small portion of which are utilized here. Our members are very willing to share more information about their experiences with the government and potential industry partners. Our members may be contacted directly or through the Data Transparency Coalition.

The projects and initiatives introduced in this section include examples of collaborative standards governance, financial reporting regimes adopting common data formats, a state government adopting a common data element, and public open data platforms.

Four Examples of Collaborative Data Standards Governance

The first example of standards governance is the **National Information Exchange Network (NIEM)** (www.niem.gov), a program housed within the Department of Homeland Security and managed by DHS, the Department of Health and Human Services, and the Department of Justice. NIEM includes both a core data element taxonomy and a development lifecycle for the creation of purpose-built information exchanges. The elements in the core dictionary and the steps of the lifecycle are available for agencies to reuse as they create information exchanges.

NEIM started the as Global Justice Information Sharing Initiative, which focused on the exchange of justice and law enforcement information across state and city government boundaries. The original initiative developed into the Global Justice XML Data Model (GJXDM), which then took on an additional homeland security focus and became NIEM.

NIEM has continued to expand in scope and participation. Beyond justice and homeland security, it has absorbed additional domains, including emergency management, immigration, maritime, and military operations. Each domain maintains its own domain-specific set of data elements and also reuses the central core taxonomy. NIEM does not yet include a formal domain related to any of the spending-related reporting regimes: federal finance, budgeting, grants, or contracts.

NIEM's data dictionary and exchange development lifecycle are formally governed through releases, with the latest being version 3.0, released in 2013.

Our second example of governance is **Standard Business Reporting (SBR)** in the Netherlands (www.sbr-nl.nl/english). Multiple Dutch agencies have agreed to use consistent data standards for the information they collect from Dutch companies.

The SBR program is administered by the Tax and Customs Administration and currently accommodates the exchange and processing of tax filings, statistics reports, and financial reports on a large scale at a national level. (SBR relies on the XBRL format, which is discussed further in the example of SEC reporting).

The SBR data standards are built into business software used in the Netherlands. Information routinely collected by organizations in the course of their business is automatically labeled according to the standards in a way that makes the data immediately ready for reporting to multiple different government agencies that use the same standards in their systems.

While today reporting is voluntary, in 2015 the first phase of mandatory filing takes place with larger companies, and all companies will have mandatory XBRL filing by January 2017. The Netherlands annually updates its SBR Taxonomy to improve on the usefulness of the XBRL data and to address changes that must comply with new legislation.

Our third example of data standards governance relates to acquisition regulations and may resonate with the procurement professionals in the contracts domain: the **Federal Acquisition Regulations (FAR) Council**, Defense Acquisition Regulations Council (DAR Council) and Civilian Agency Acquisition Council. These councils routinely bring together federal agencies to maintain a set of rules – including some data standards – that apply across the government.

Our fourth example is the **Open Contracting Partnership** (www.open-contracting.org), a global organization housed within the World Wide Web foundation and funded by the World Bank and the Omidyar Network. The Open Contracting Foundation to create and promote both a data standard and a set of best practices for public-sector procurement.

The Open Contracting Partnership released the first version of its Open Contracting Data Standard, which is freely available for reuse, on November 18, 2014.

Three Examples of Financial Reporting Regimes Adopting Data Formats

The first example of a financial reporting regime adopting a data format is the **Securities and Exchange Commission (SEC)** (www.sec.gov). In 2009, the SEC adopted the XBRL format for the financial statements submitted by public companies under the securities laws. XBRL assigns a unique electronic tag to each item in a financial statement. The agency has also adopted XML formats for a few other forms that it collects under the securities laws.

The SEC has continued to collect a plain-text version of each financial statement alongside the XBRL version, and until July 2014 did not enforce the quality of the XBRL version, the benefits of the SEC's adoption of XBRL have been limited. Moreover, the SEC's adoption of data standards has been incomplete; nearly all of the agency's hundreds of forms are still expressed as documents, not as standardized data. However, the potential remains for the SEC to dramatically improve the accountability of the U.S. capital markets to investors, facilitate data analytics to illuminate potential violations, and allow its registrants to automate compliance. To these ends, the Data Transparency Coalition has advocated for the SEC to fully enforce its existing data standards and to embrace a complete transformation for all of the forms that are still expressed as documents.

The second example of a financial reporting regime adopting a data format is the **Federal Financial Institutions Examination Council (FFIEC)** (www.ffiec.gov), which is made up of the Federal Deposit Insurance Corporation, the Federal Reserve System, and the Office of the Comptroller of the Currency. The FFIEC adopted the XBRL format for banks' call reports in 2005, reducing error rates from approximately one-third to nearly zero and reducing preparation time from weeks to days.

The third example is the **European Union's Solvency II** project, which aims to codify and harmonize the European prudential regime for insurance and reinsurance. Solvency II covers more than 18,000 companies regulated by the European Union Insurance and Occupational Pension Authority (EIOPA). EIOPA worked collectively across multi-state regulators and insurance agencies to define and harmonize the data they collect. EIOPA, too, utilized the XBRL format to standardize reporting.

Example of State Government Adopting a Common Data Element

One early example of the value of common data elements is the State of Michigan's adoption of a common data element to identify the recipients of state funds. By identifying recipients consistently across more than twenty reporting regimes, Michigan realized a savings of over $1 million per business day in fraud reduction. As part of the same project, the state consolidated 40 data centers into three, saving nearly $30 million.

Open Data Publication Platforms

Across the United States, state and local governments have established publication platforms that publish their spending information more comprehensively than the federal government does. Although many state and local governments have yet to adopt consistent data standards and face similar challenges to the federal ones discussed above, several have managed to connect concepts across multiple reporting regimes. Useful examples include the City of Boston, Massachusetts, which launched Open Budget, and Montgomery County, Maryland, which enlarged an existing open data program to make its budget and spending data more accessible.

V. Examples as Illustrations in Response to Treasury's Questions

Treasury has requested examples in respect of specific criteria. This section answers Treasury's questions with references to the foregoing discussion. Our responses in Section V include recommendations and lessons learned.

A. "Open"

Treasury asked for a description or examples of data standards on data exchange that could ensure the data is "open." Treasury defined "open" in this context as meaning that "anyone can access, use, or re-use posted information, including the public, Federal agencies, local and state governments, academia, media, industry, standard-setting bodies, transparency groups, on a worldwide scale."

1. Nonproprietary data standards. To ensure that federal spending data is "open," Treasury and OMB must adopt nonproprietary data standards (see Section II). If standardized data elements or a format are subject to licensing or any other reuse restrictions, Treasury's first criterion simply cannot be met.

If they seek to make federal spending data "open," under this definition, Treasury and OMB must replace the existing proprietary DUNS number with a nonproprietary identifier.

2. FFIEC example. The effective "access, use, or reuse" of information requires that information to be expressed as standardized data, not as documents.

A prime example of the replacement of documents with standardized data is the FFIEC's adoption of XBRL as the format for data collection from banks (see section IV). Before XBRL adoption in October 2005, 8,200 banks were each quarter reporting an average of 25 million data points, including 16,000 to 18,000 errors (*i.e.*, 2 to 3 per bank). These errors required federal analysts to call the banks, get the errors corrected, and ask for the data to be resubmitted. After resubmission, the data was published within 60 days on the FFIEC's Uniform Bank Performance Report (UBPR). XBRL validation before submission by each bank reduced drastically the number of errors, allowing the FFEIC to publish its performance report only a few days after submission. The FFIEC's improved

reuse of banks' data shows the value of data standards to fulfill Treasury's "open" criterion.

3. SEC counter-example. Because the SEC has continued to collect duplicative documents alongside its XBRL-formatted financial statements (see Section IV), and because the agency has not yet fully enforced the quality of the XBRL-formatted financial statements, the full benefits of openness have not yet been realized. Investors and the information providers serving them have been slow to adopt XBRL-based analysis tools.

4. Embedded tools alongside bulk download. It is important to understand the audiences and consumers of open data. Some audiences may be most interested in simple visualizations provided directly by the government sources of such data. Others may want direct access to the raw data so that they can apply more sophisticated tools. To make federal spending data fully "open," Treasury and OMB should ensure that it is presented alongside effective basic analysis tools while also making it fully downloadable in bulk.

For example, the City of Boston, which has an open data site that offers raw datasets for download (www.data.cityofboston.gov), also has an Open Budget platform to provide Bostonians with a more unique experience navigating the operating and capital budgets datasets being posted.

B. "Availability"

Treasury has defined availability in this context as "free access to the data standard, both during development, at final stage, and for translations." According to Treasury, availability is also "assurance that core technologies can be implemented royalty-free."

1. Nonproprietary data standards. By definition, proprietary data standards cannot be accessed freely. Treasury and OMB must adopt nonproprietary data standards (see Section II).

2. Sustainable governance. At the outset federal government should manage or steward the data standards using a mature and predictable governance structure (see Section II). The governance structure should make it easy for non-federal stakeholders to give substantive technical input and, especially, to access all relevant technical documentation.

We believe that after DATA Act data standards have been established and applied, Treasury and OMB should consider the option of transferring governance to a non-federal entity for the long term.

3. NIEM example. NIEM is an open standard for specifying "Information Exchange Package Documents" (IEPDs), which are used by federal and non-federal entities to govern specific information exchanges (see Section IV). Defined and registered IEPDS

may be reviewed and reused by anyone because they are available in the NIEM clearinghouse.

NEIM governance and implementation is coordinated by an executive steering council and a project management office. These structures support a NIEM Technical Architecture Committee and a NIEM Business Architecture Committee, which are in turn complemented by committees focused on specific domains. Significantly, the NIEM model, training, information about NEIM communities, and even an "implementation cookbook" are all freely available NIEM's website.

4. Reuse of widely-accepted formats. To maximize the availability of any standards, Treasury and OMB should reuse standards developed and maintained by others (Section II), rather than developing new ones, wherever possible. For example, if Treasury and OMB adopt XBRL as the format for federal spending, they will be adding to a well-developed global community of XBRL users and practitioners, rather than starting a new one from scratch. All of the examples related to XBRL (Section IV), which is maintained in the US by XBRL US and elsewhere by XBRL International, feature this benefit. Similarly, if Treasury and OMB adopt the Open Contracting Data Standard to govern the Contracts reporting regime, they will be able to take advantage of a global community of practice and software development.

C. "Business Reach"

Treasury asks about efficient "business reach" to foster "private sector innovation," and states that "business reach" refers to the global reach of the business community.

The adoption of government-wide data standards for federal spending will support the development of new business models in our industry.

The Coalition has focused on three benefits of data transparency in federal spending: accountability, management, and automated compliance. We believe that each of these three benefits corresponds with a business opportunity for our members.

Some Coalition members republish government data, facilitating accountability while earning revenue from advertising or subscriptions. Others offer analytics software to federal managers; these companies realize that data standards would allow their products to deliver better insights at less cost. A third category of Coalition members includes companies seeking to automate compliance burdens for grantees and contractors. Data standards for recipient reporting will allow these members to replace manual tasks with automatic processes.

In addition, there is already an international market for consulting and software solutions to develop and deliver financial information as standardized data. Our members have provided support to all of the projects discussed above (Section IV).

D. "Validation"

Treasury seeks examples of appropriate "validation" and raises the issue of whether validation of business rules should be "rules-based" or "document-based." Treasury defines rules-based as where the business rules are "interwoven into the fabric of the standard output and standard governance." Treasury defines "documents based" as being "separately stored and not entirely contained within the standard output or standard governance body."

The DATA Act's standards mandate offers an opportunity for the federal government's existing spending reporting regimes to improve the quality and efficiency through incrementally improving rules-based validation. As they establish data standards, Treasury and OMB should seek to understand the particular quality challenges of each regime and strike an appropriate balance between easily achievable rules-based validations and those that might be more difficult to automate or require agency-specific approaches.

1. FFIEC example. The FFIEC's adoption of XBRL for banks' call reports represents a move from purely document-based validation to a combination of rules-based and document-based validation. The FFIEC's system can reject call reports with unacceptable values or with mandatory fields submitted as blank.

2. SEC counter-example. The SEC's failure to deliver the promised benefits of open data in corporate financial statement reporting is associated with a failure to validate. Even after starting to collect XBRL-formatted financial statements, the SEC continued to manually check their compliance with mathematical rules, rather than using software to check calculations. The SEC retained an almost purely document-based validation approach despite collecting sufficient structured data to support some rules-based validation.

3. Selecting a data format that is supportive of rules-based validation. Some data formats are more supportive of validation than others. For example, XBRL allows users to define, express, and check the relationships between data elements, while simpler data formats such as CSV do not.

4. NIEM example. NIEM does not centrally maintain validation rules, but some organizations using NIEM, including the Department of Justice, have developed their own tools to help validate NIEM-formatted data. In a similar way, agencies administering reporting regimes using DATA Act data standards might choose to adopt their own rules-based validations to improve the data quality of submissions.

E. "Extensibility"

Treasury refers to "extensibility" as an issue of maintaining "future modifications to the data standard" and describes a spectrum running between flexible and rigid. Treasury also says that flexible here means "implementation ease."

1. Incremental implementation requires variable extensibility. As Treasury and OMB establish data standards for federal spending, they should take into account the needs of each reporting regime to which the standards will apply. Some existing federal spending reporting regimes will benefit from greater flexibility through extensibility; others will require more rigidity.

For example, federal payment requests are submitted using a relatively short list of data elements. Treasury's systems require rigid adherence to these data elements in order to process payments. Treasury and OMB might choose to maintain the rigidity of the elements already being used for federal payment requests. On the other hand, federal account balance reporting involves more complex line items and must accommodate some agency-specific concepts. Treasury and OMB might choose to allow some extensibility for this reporting regime.

The Data Transparency Coalition recommends rigidity in the DATA Act data standards' governance and the issuance of future versions, but also recommends that each version of the standard allow some well-defined flexibility based on the needs of specific reporting regimes.

2. NIEM example. NIEM handles many disparate domains, from health care to homeland security, with varying needs for extensibility. The NIEM program offers two types of model updates: major and minor releases. Major releases occur when the NIEM core dictionary and domains are updated and then synchronized. Additionally, any technical architecture changes to the model are exclusively made during a major release. Major releases are given version identifications s such as 2.0, 3.0, or 4.0. Since 2005, NIEM has issued four major releases: 1.0 in 2006, 2.0 in 2007, 2.0 in 2009, and 3.0 in 2013.

In contrast, minor releases occur to incorporate and synchronize changes to domain-specific content only. The NIEM core dictionary does not change in a minor release. Only those updates that can be applied to domains or code lists are integrated. Minor releases are given version identifications such as 2.1, 2.2, 3.1, or 3.2.

Treasury and OMB could choose to maintain DATA Act data standards in a similar manner. Top-level or core updates could be periodical, with more frequent updates of domain specific supplements. These could be referred to as major and minor releases.

3. SEC counter-example. The SEC's XBRL implementation includes a standard set of nearly 20,000 tags that represent U.S. GAAP. Additionally, however, public companies have broad flexibility to use a their own custom tags for financial elements within a balance sheet, income statement, or statement of cash flow. There has been wide spread use of custom tags for many items that are not unique and this has made it difficult for investors, analysts, and financial market researchers to make inter-company comparisons of the financial disclosures.

F. "Ease of Implementation"

The Treasury's definition of ease of implementation refers to "integration and interoperability within a given environment." The Treasury sees difficult implementation where "[an] integration environment necessitates extra steps."

The Data Transparency Coalition believes the implementation of data standards for federal spending will ultimately reduce the number of steps required in federal spending reporting, ease data exchange, and reduce transaction costs.

G. Benefit to Constituents and Stakeholders

Treasury asks how the examples or descriptions provided in A-F above would, if implemented, benefit or add value to our constituent group or pertinent stakeholders.

1. Direct benefits. The Data Transparency Coalition is the nation's only open data trade association. The Coalition represents leading technology and management consulting firms, including industry leaders and growing startups, some offering specific solutions and others whose services are solution-agnostic. If Treasury and OMB implement the recommendations in these comments, our member companies will be able to help their clients, customers, consumers, or users achieve better accountability, better management, and automatic compliance in federal spending reports.

Aside from the Coalition's corporate members, we believe Treasury and OMB should consider the following constituencies and direct benefits:

- **Enhanced Accountability.** Citizens, media, and watchdog groups will enjoy better access to the information necessary to follow the federal government's use of its resources.
- **Better Management.** Congress, executive branch leaders, inspectors general, agency heads, financial managers, grant and contract officers, and program managers will benefit from data-driven decision-making capability and related tools.
- **Automated Compliance.** Recipients of contract, grant, and other funding will be able to replace manual compliance tasks with automated processes. **[amplify]**

2. Indirect benefits. Beyond the direct benefits of standardized federal spending data, the adoption of the data standards mandated by the DATA Act might result in the following indirect benefits as well.

- **Reuse of data standards.** Once DATA Act data standards are have been established, assuming that they are fully nonproprietary, they will be freely available for reuse within reporting regimes not directly related to federal financial, payment, budget, contract, grant, and recipient reporting. For example,

OMB might choose to use DATA Act elements to automate some aspects of compliance with the Single Audit Act.

• **Connecting federal spending data with data from other reporting regimes.** The standardization of federal spending data will facilitate connections between federal financial, payment, budget, contract, grant, and recipient data sets and others beyond spending. For example, if Treasury and OMB adopt the Legal Entity Identifier (LEI) to identify grantees and contractors and U.S. financial regulators also adopt the LEI for companies reporting to them, investors could automatically connect each company's financial regulatory filings with its government contracting records.

H. Anticipated/Envisioned Use Cases

Treasury is seeking example use cases that we anticipate or envision for information with data structured in accordance with established data standards.

We believe that the ultimate success of the DATA Act should be judged by whether, and when, the following use cases become viable.

1. Full life-cycle transparency. Treasury has envisioned that data standards could link together all of the sequential stages of federal spending information, from Congressional appropriation to Treasury allocation to agency obligation and expenditure to final award and subaward. Congress, Presidential administrations, agencies, and recipients – not to mention citizens and media – would all benefit from the ability to track the ultimate result of a particular appropriations line item, in terms of dollars actually spent.

2. Geospatial views of internal and external spending. Robust geospatial standards for both (1) external grants, assistance, and contracts and (2) internal spending, such as on salaries and supplies, will allow the geographic impact of all federal spending to be aggregated. With full geospatial standards, fully applied, the former Consolidated Federal Funds Report, canceled in 2010 by the Census Bureau due to the expense required for its manual preparation, could be replicated and instantly modified.

3. Interactive views of entity hierarchies. Divisions and subdivisions of all entities that interact with federal spending – both federal agencies and nonfederal grantees and contractors – should be consistently tracked and identified. Users should be able to view each of the units of an agency or company that collects or receives federal funds and understand each unit's activities.

4. Automatic grantee and contractor reporting. Grantees and contractors should be able to use software to automatically comply with most federal reporting requirements.

I. Impact On Constituents and Stakeholders

Treasury is seeking input on the impact that data standards on data exchange would have on our constituent group or pertinent stakeholders, and investments.

The Data Transparency Coalition sees only positive impact for our members and their investments. (We refer to Subsection G above, regarding benefits.)

However, we note again that not everybody will welcome the efficiency or innovation promised by the Data Act. Change can be disruptive to those who have built a business model that is outdated or have sunk investments into complying with inefficient processes. Progress is often resisted by those who could lose market share to new competition as the barriers to market entry are reduced through simplification.

J. Other Criteria

The Treasury invites suggestions of other criteria that might be considered by Treasury and OMB in establishing the data standards on data exchange.

As stated in Section II, the Data Transparency Coalition believes that the data standards, their implementation, and the resulting standardized data must fulfill certain essential criteria to realize the promised benefits of the DATA Act. First, the data standards themselves itself must be *complete*, *accepted*, and *nonproprietary*. Second, the implementation of the standard must be *enforced, incremental*, and *sustainable*. Third, the data that is exchanged through the system must be *supportive of validation*. (See Section II.)

VI. Conclusion

The Data Transparency Coalition advocates the publication of government information as standardized, machine-readable data. We believe that open data will facilitate public accountability, enables data-driven government management, and automates compliance – in federal spending under the DATA Act, and in other types of federal reporting under future mandates.

With full implementation, the DATA Act will unlock a new public resource that innovators, watchdogs, and citizens can mine for valuable and unprecedented insight into federal spending. The DATA Act could deliver cost-savings in government by reducing red tape and allowing information to flow better between silos. The DATA Act has the potential to modernize government, giving stakeholders the tools to understand, monitor, predict, and make decisions using timely data on actual expenses – just as America's leading corporations do already. Finally, the DATA Act can relieve the burden of compliance through automated reporting.

The Data Transparency Coalition is grateful for the opportunity to submit comments and is eager to expand upon any of the issues raised in this paper. We thank Treasury and OMB for their ongoing efforts.

November 25, 2014
Washington, DC.

This paper was prepared by J. Ruairi Macdonald who was on Field Placement with the Data Transparency Coalition as part of a Master Degree (LL.M.) in Government Procurement Law program at the George Washington University Law School. Mr. Macdonald was supervised by Data Transparency Coalition Executive Director Hudson Hollister.