Testimony

of Ellen Miller
Co-founder and Executive Director
The Sunlight Foundation

before the Committee on Oversight and Government Reform
Subcommittee on Technology, Information Policy,
Intergovernmental Relations, and Procurement Reform
United States House of Representatives

on Transparency through Technology:
Evaluating Federal Open Government Efforts

March 11, 2011

Comments of the Sunlight Foundation

Before the

Committee on Oversight and Government Reform
Subcommittee on Technology, Information Policy, Intergovernmental Relations, and Procurement Reform

March 11, 2011

Chairman Lankford, Ranking Member Connolly, and members of the Committee, thank you for the invitation to appear before you today.

My name is Ellen Miller and I am the co-founder and executive director of the Sunlight Foundation, a non-partisan non-profit dedicated to using the power of the Internet to catalyze greater government openness and transparency. We take inspiration from Justice Brandeis' famous adage "Sunlight is said to be the best of disinfectants."

We believe that the public has a right to know how its government works. I am pleased to have the opportunity to speak to you today about the progress that has been made toward achieving that goal. Recent Congresses and presidential administrations deserve congratulations for taking concrete steps toward embracing a 21st century vision of transparency. Initiatives like the Open Government Directive are emblematic of a willingness to take the issue seriously.

Unfortunately, the OGD's value has proven to be primarily aspirational. While establishing positive transparency norms is important, we believe that the government must rededicate itself to addressing the fundamental challenge of transparency itself. Merely acknowledging transparency's importance is no longer sufficient; transparency initiatives must now yield information that is accurate, complete and useful.

The Data Powering USASpending.gov Are Seriously Flawed

There is perhaps no better example of this tension than USASpending.gov.[1] Disclosure of the ways in which the public's money is spent is among the most important types of government transparency. Congress acted upon this responsibility in 2006 with the passage of the Federal Funding Accountability and Transparency Act,[2] which required that information about federal grants, contracts, loans and insurance be placed online in a searchable website known as USASpending.gov.

In the course of their work, Sunlight researchers have become deeply familiar with the data powering USASpending.gov. As we began to examine these systems, we were aware the quality of these datasets was widely considered to be problematic, based both upon GAO reports[3] and the informal consensus of the research community. Our work quickly confirmed that the data suffered from irregularities. We were anxious to reach a more complete understanding of the problem.

In order to do so, we needed a reference point against which we could compare USASpending.govdata. Unfortunately, the complexities of federal budgeting make both budget and U.S. Treasury expenditure data unsuitable for this use. However, we found

---

[1] The remainder of my testimony will focus on financial spending transparency. For my analysis of the Open Government Directive, please see my speech at the Gov2.0 summit, available here: http://bit.ly/gMli7P, and my blogpost "Open Government: idling in the driveway," available here: http://bit.ly/gbb76O.

[2] P.L. 109-282, available at http://1.usa.gov/eWVqSR.

[3] See, e.g., "Nonprofit Sector: Significant Federal Funds Reach the Sector through Various Mechanisms, but More Complete and Reliable Funding Data Are Needed" (GAO-09-193) (available at http://www.gao.gov/products/GAO-09-193); "Electronic Government: Implementation of the Federal Funding Accountability and Transparency Act of 2006" (GAO-10-365) (available at http://www.gao.gov/products/GAO-10-365).

our yardstick in the Catalog of Federal Domestic Assistance,[4] an index of many federal programs that includes program descriptions and yearly obligation amounts. Although not strictly designed for this use, the CFDA has been used for comparative analysis by GAO.[5] We took GAO's methodology and expanded it. We looked for mismatches in amounts between the CFDA and USASpending.gov, allowing a generous margin of error[6] to account for differences between the systems. We also looked for instances in which reports had not been made within statutory deadlines, and for incomplete reports. Finally, we automated GAO's sample-based methodology so that we could examine the entire database, which consists of hundreds of thousands of records.

The results were sobering. We found **over 1.2 trillion dollars' worth of misreported spending in 2009 alone**.[7] Some of the most serious problems appear to be caused by agencies' failure to meet their reporting obligations. The USDA website lists the cost of their school breakfast and lunch programs at $12.7 billion, but only $250,000 of these costs are reported on USASpending.gov.[8] The Maritime Administration has never reported the spending associated with any of its loan or insurance programs, and

---

[4] Available at https://www.cfda.gov/

[5] See, e.g., "Electronic Government: Implementation of the Federal Funding Accountability and Transparency Act of 2006" (GAO-10-365) (available at http://www.gao.gov/products/GAO-10-365).

[6] Our analysis ignores disparities between the CFDA and USASpending.govthat are less than fifty percent of the total dollar amount for the program in question.

[7] http://sunlightfoundation.com/clearspending/

[8] See http://1.usa.gov/hMl51A (school breakfast FY09) and http://1.usa.gov/gNpmbM (school lunch FY09).

reports only a fraction of its grant activity. These are just two examples. Almost every agency has one or more programs that fail to report their spending.[9]

And the spending that is reported is often incomplete or incorrect. For example, each loan record is required to include both the subsidy cost and face value of the loan.[10] Unfortunately, the subsidy cost is incorrectly reported as zero for over 85% of all loan records. And the face value of all FY2010 student loans was report as $6.9 trillion – an amount greater than the entire federal budget. This is clearly wrong.[11]

Ironically, agencies typically use purpose-built internal systems for managing their spending that are separate from these public reporting systems and much more accurate. In essence, agencies are maintaining two sets of books. Publicly disclosing relevant information from these internal systems is vital both for the public and for government planning efforts like the Consolidated Federal Funds Report.[12]

We recognize that the Catalog of Federal Domestic Assistance may have its own reporting errors, but we think it is fair to draw attention to the gross inconsistencies between CFDA and USASpending.gov spending data.[13] The $1.2 trillion mis-reporting number does not include contracting data, as the CFDA does not provide a point of

---

[9] http://sunlightfoundation.com/clearspending/scorecard/

[10] In other words, an estimate of the true cost of the loan to government, reflecting the risk of default, loss of interest revenue, and other factors.

[11] The problems with loan reporting are so extensive that we did not include them when we calculated the $1.2 trillion discrepancy between CFDA and USASpending.gov.

[12] CFFR is a presentation of data on federal government expenditures or obligations in state, county, and subcounty areas of the United States, including the District of Columbia and U.S. Outlying Areas, available at http://www.census.gov/govs/cffr/.

[13] We attempted to manually correct order-of-magnitude errors in the CFDA, such as when billions of dollars were inadvertently substituted for millions.

comparison for procurement. And, the number also does not reflect the myriad of problems with loan data, which we discussed above.

The Challenge of Useful Identifiers

The data are not only limited by these failures of consistency, timeliness and completeness. Another major limitation is that the identifiers used throughout the system are poorly designed, unreliable, and act as barriers to use of the data. Identifiers are the license plates of data, and their unreliability creates confusion.

Within the dataset, identifiers are used to designate unique entities. A single recipient of multiple contracts should be assigned a single identifier to ease the examination of those contracts as a set. For example, if Boeing wins several contracts to build a number of planes, Boeing itself should have a unique identifier. Similarly, all of the payments associated with a grant should have a unifying grant ID. For example, if a homeless shelter receives monthly payments for the services it provides, all of the money should be identified as belonging to a particular grant. Finally, individual payment records should have unique identifiers. In other words, each transaction must be identified, like checks in a checkbook. Consequently, different databases tracking the same information will have the same data, like when reconciling a checkbook to a bank statement.

These are all obvious and uncontroversial requirements of a well-designed database system. Unfortunately, the federal government's spending systems fail to satisfy any of them.

Perhaps the most challenging problem involves recipient identifiers – identifiers that indicate who has received money. The government spends millions per year on purchasing a private sector solution to identifying recipients of federal grants and contract dollars from Dun & Bradstreet.[14] Obtaining a DUNS number is a requirement for receiving many kinds of federal awards.

Unfortunately, using DUNS numbers carries severe limitations. Business entities often have complex structures, and it is necessary for a single entity to possess many identifiers. The linkages between these identifiers are as important as the IDs themselves. Dun & Bradstreet maintains a vast system of these linkages. Unfortunately, **none of them are freely available to the public** as it tries to make sense of data obtained from USASpending.[15] Only the lowest, most granular level of DUNS numbers is exposed. This is not adequate.

For instance, each individual Wal-Mart location has a unique DUNS. Without the ability to link these IDs to the parent corporation -- to identify the branches of the same tree -- an analysis of funds flowing to Wal-Mart will be incomplete. Various parts of the federal government already collect the information that could replicate the functions of DUNS.[16] But transitioning from the existing system will be a substantial undertaking, as it poses challenges related to system upgrades, privacy, and political sensitivities.

---

[14] The total amount spent on Dun & Bradstreet subscriptions is unclear because the contract reporting lacks specificity. For the total amount spent on Dun & Bradstreet, see this spreadsheet taken from USASpending.gov data: http://bit.ly/hSNdFw. In FY2010, it was approximately $56 million.

[15] And even once access to the Dun & Bradstreet database is purchased, because of its proprietary nature, the publication of linkage data and any results that could reveal those linkages is prohibited. As a result, public oversight is dramatically curtailed.

[16] For background information, please see http://bit.ly/fFT7jI

Nevertheless, because data transparency should be an essential part of government's mission, we consider the government's adoption of a truly open identifier system to be a vitally important undertaking.

The problems of award and record identifiers are considerably more manageable. Award IDs suffer from a simple lack of coordination: each agency generates numbers in an ad-hoc manner, without any central structure or guidance. It should be relatively simple to coordinate the generation and assignation of award IDs. Last year the Recovery Accountability and Transparency Board issued a whitepaper entitled *Solutions for Accountability and Transparency: Uniform Governmentwide Award ID Number*[17] that discussed this problem at length. We endorse both its diagnosis of the problem and recommendations. Issues surrounding record identifiers are the simplest problem to resolve.[18]

The need for reliable IDs is clear, and the path to satisfying that need is straightforward.

Conclusion

We do not believe that the problems affecting USASpending.gov are the fault of the website or the people that maintain it. Indeed, USASpending.gov deserves praise for its growth and improvement: when we first conducted this analysis, we had to obtain a

---

[17] http://1.usa.gov/9pkpzq

[18] Indeed, until recently record identifiers have been quite reliable and useful to us in our analysis. A major overhaul of the record IDs present in the USASpending.govdata posed substantial analytic challenges for Sunlight. We believe that this change was made in good faith, likely due to a belief that no one outside of the engineers maintaining USASpending.gov utilized these IDs. This is understandable, but speaks to the lack of attention and rigor that government currently devotes to the identifier problem.

copy of the data by shipping a hard drive to Maryland. Today we are able to download it directly from the USASpending.gov website. Similarly, we were pleased to see the administration finally begin to offer the subaward data mandated by FFATA.[19] But these improvements will be meaningless for the vast majority of users unless the underlying data can be made reliable. Until agencies begin to take this responsibility more seriously, federal spending transparency will remain an unfulfilled promise. And what is true for federal spending transparency is equally true for the Open Government Directive.

We welcome the Committee's attention to this issue and encourage you to continue to spend time engaging in oversight and legislative efforts. Thank you for the opportunity to testify today. I look forward to your questions.

---

[19] The statutory deadline for this data's availability was January 1, 2009. Responsibility for the lateness of the data properly rests with many parties.

Ellen Miller

Ellen S. Miller is the co-founder and executive director of the Sunlight Foundation, a Washington-based, non-partisan non-profit dedicated to using the power of the Internet to catalyze greater government openness and transparency. She is the founder of two other prominent Washington-based organizations in the field of money and politics -- the Center for Responsive Politics and Public Campaign -- and a nationally recognized expert on transparency and the influence of money in politics.

Her experience as a Washington advocate for more than 35 years spans the worlds of non-profit advocacy, grassroots activism and journalism. Ms. Miller's work has been featured in Government Technology ("Top 25 Doers, Dreamers & Drivers in Public Sector Innovation," March 2011), Washingtonian Magazine ("100 Tech Titans," May, 2009), Fast Company, ("The Most Influential Women in Technology," January, 2009), WIRED Magazine ("15 People The Next President Should Listen To," October, 2008), The Chronicle of Philanthropy ("Seeking Online Exposure," January, 2008).

Ms. Miller also served as deputy director of Campaign for America's Future, was the publisher of TomPaine.com and was a senior fellow at The American Prospect. She spent nearly a decade working on Capitol Hill.

Committee on Oversight and Government Reform
Witness Disclosure Requirement – "Truth in Testimony"
Required by House Rule XI, Clause 2(g)(5)

Name: Ellen Miller

1. Please list any federal grants or contracts (including subgrants or subcontracts) you have received since October 1, 2008. Include the source and amount of each grant or contract.

None

2. Please list any entity you are testifying on behalf of and briefly describe your relationship with these entities.

Co-founder and executive director of the Sunlight Foundation.

3. Please list any federal grants or contracts (including subgrants or subcontracts) received since October 1, 2008, by the entity(ies) you listed above. Include the source and amount of each grant or contract.

None

I certify that the above information is true and correct.

Signature: Ellen S. Miller

Date: March 1, 2011