MERCATUS CENTER
GEORGE MASON UNIVERSITY

Testimony of
Eileen Norcross, M.A.
Senior Research Fellow
Mercatus Center at George Mason University

Before the
House Committee on Science and Technology

March 19, 2009

Mr. Chairman and Members of the Committee:

Thank you for inviting me to testify on "Follow the Money: Accountability and Transparency in Recovery Act Science Funding." My research at Mercatus has included considering ways to improve the access, clarity, and usability of government data so agencies, Congress, and, researchers and citizens, can make better use of that data, while helping ensure greater accountability for government spending.

The Committee knows how vital transparency and accountability is to the progress and performance of the nearly $800 billion in funds committed as part of the American Recovery and Reinvestment Act (ARRA).

Following the stimulus money—as it branches out from agencies to state, and local budgets, contractors, subcontractors, individual projects and transactions is a monumental and frankly, impossible task for a centralized entity, no matter how many auditors and analysts the government commits to the job. Simply put: information about how funding is ultimately spent is dispersed, and knowledge about how funds are used is local. A central entity cannot possibly marshal or even be alert to all of the possible transactions and dedications of stimulus dollars, or comprehensively assess how projects are performing.

Those in government charged with this important task are being asked to manage an unprecedented deluge of data that accompanies a funding commitment of this size. There is the added pressure and paradox of spending funds quickly, while ensuring prudent management. As the Professional Services Council stated, "without a government workforce sufficient to plan, deliver, and manage the contracts and grants that dispense these huge funds, it will be like constructing an office building on a foundation of sand."

Fortunately that workforce can be augmented. The public—individuals with local knowledge—aided by technology, can fill in the gap. To help government with the stimulus-monitoring effort, there are low-cost innovative solutions that embrace the best principles of government accountability, 21st century Internet technology, citizen, and community involvement. I would like to highlight one such approach today—crowdsourcing.

Crowdsourcing or "peer production" allows a large group of people to make small individual contributions to a project or product that would traditionally have been produced by a single individual or organization.[1]

The most familiar example of crowdsourcing is Wikipedia—the online community-oriented encyclopedia. A wiki is a kind of online collaborative notebook. Wikipedia is a reference built by volunteers who add, expand, enrich and edit reference articles on any topic conceivable.

The result is a vast, ever-evolving, but easy-to-use resource that is more extensive than what a limited number of editors could produce. Importantly, Wikipedia is dynamic, continually updated and monitored by users for content and accuracy.

This kind of informational dynamism, which permits ongoing content and data enrichment and improvement, has only become possible with Internet technology. By reducing the transaction costs between individuals, the Internet and continual software developments allow people to gather and analyze data in novel and creative ways—to take apart seemingly impenetrable masses of data and extract patterns and new meaning.

But it is only possible if government provides data according to a few basic principles, advanced by my colleague Jerry Brito.

Data must be structured, open, and searchable. In other words, it must be provided in useful formats, standard, web-friendly, machine-readable formats that can be aggregated, parsed, and sorted. A loose analogy is to think of this as data in a spreadsheet–with rows and columns that allow users to sort according to criteria and uncover trends and patterns.

This is in contrast to disclosing spending in reporting narrative. The information might all be there, but it doesn't allow a computer to analyze it. However, putting information in XML (Extensible Markup Language) would allow a user to search, for instance, all projects over $500,000 for a contractor in a particular state, or congressional district.

Providing data in this form allows users to innovate, building tools to analyze and improve upon the data, merging maps, economic statistics, industry information thereby enhancing reporting. This informs citizens, and also aids the government in the effort to gauge how federal funds are performing.

I will now turn to a concrete example of how this is working in practice.

<u>What crowdsourcing can do for stimulus accountability</u>

---

[1] Jerry Brito, *Hack, Mash & Peer: Crowdsourcing Government Transparency,* 9 Columbia Science and Technology Law Review 119 (2008), available at http://www.stlr.org/html/volume9/brito.pdf

One of the great benefits of crowdsourcing is that it is low-cost and fast. The human capital is already in place – made up of volunteer programmers and good government activists in the online transparency community.

1) *StimulusWatch.org*

In early December I came across the online U.S. Conference of Mayors Ready-to-Go wish list of projects cities submitted detailing how they would like to spend potential stimulus funds.

The list had several virtues. Importantly it provided granular details of how federal funds might be spent on the local level. It was possible to sort by city, and federal funding type. Each project listed included valuable details: project dollar amounts, potential jobs created, and in most cases, brief project descriptions.

However, there was a limited amount of information I could extract by myself not having local knowledge of all of the communities nor unlimited time. The data was meaningfully displayed, but the format was relatively rigid and did not allow for keyword searching.

I asked my colleague Jerry Brito if the list might be a good opportunity for crowdsourcing—to invite people with local knowledge to improve the content while also proving localities with feedback on the relative merit of individual projects.

Jerry with the help of two volunteer software developers, Kevin Dwyer and Peter Snyder screen-scraped the data from the Mayor's site and within a few weeks created StimulusWatch.org. The site uses the data reported by participating cities to the US Conference of Mayors, while improving upon its usability by allowing visitors to vote and comment on individual projects and search projects by keyword.

One of the most important features of the site is that it allows individuals—citizens and city officials alike—to contribute wikis, or factual information, on individual projects, in many cases enhancing and clarifying the project descriptions initially provided by individual cities. In effect, StimulusWatch.org helped transform a static report into a kind of online national Town Hall.

We observed some interesting trends. First, the response was tremendous. In the first month we had two million unique visitors, and many journalists using the site. Users were actively commenting on projects and adding information.

Projects deemed low-priority – dog parks, for example, rose to the top within hours of the site's launch. An official in Natchez, Mississippi clarified in the local paper a few weeks after the site's launch that a nature trail project listed on the site as costing $600 million, was only supposed to cost $3.1 million – an error had been made, and it had been caught, because so many eyes were able to quickly parse the data.[2]

[2] "City did not ask for $600M for project" by Mary Hood *Natchez Democrat*, February 7, 2009, http://www.natchezdemocrat.com/news/2009/feb/07/city-did-not-ask-600m-project/

Other projects, in particular in transit and light rail, quickly emerged as high-priority, stimulating active and ongoing debate by local citizens. Many project descriptions were enhanced by people contributing factual information, in some cases, clarifying what were initially regarded as wasteful projects.

The result of StimulusWatch.org to date has been more than encouraging, it has functioned as a demonstration project—showing how the public can contribute and help the government keeps its commitment to accountability by ferreting out potential waste, while also becoming civically engaged, providing feedback to officials on how dollars are ultimately used to benefit the community.

2) *Crowdsourcing and Stimulus Accountability*

The online community is eager to help government in the task of monitoring stimulus funding. Recovery.gov has made a commitment to provide information to the public.

In order to meet the government's goals of transparency and accountability, certain details must be provided to the public. Specifically,

a) *Project-level details on how funds were spent.*
Individual should be able to drill down from contractor, to subcontractor, to the level of individual transaction, up to a cap of $25,000.

b) *Absent government provision of a database, raw data should be made available.*
The government does not have to build such a database to track spending, it only needs to require that grantees (states, localities, and grantees) provide raw data according to the principles mentioned—structured, open, and searchable, and then make that data available to the public. This will allow users to access, search, and analyze data for patterns and trends.

But before this is possible, the disclosure and transparency requirements in the American Recovery and Reinvestment Act must be strengthened. ARRA does not require data be provided in structured machine-readable formats. Guidance issued by the Office of Management and Budget does not remedy other outstanding issues.

The Recovery Accountability and Transparency Board must address four issues in how it requires and will publish data: detail in disclosure, standardization of information, aggregation and centralized access.

a) *Detail and Depth in Disclosure*
According to OMB's guidance,

"Reporting requirements only apply to the prime non-Federal recipients of Federal funding, and the subawards, (i.e., subgrants, subcontracts, etc.) made by these prime recipients. They do not require each subsequent subrecipient to also report. For instance, a grant could be given from the Federal government to State A, which then gives a subgrant to City B (within State A), which hires a contractor to construct a bridge, which then hires a subcontractor to supply the concrete. In this case, State A is the prime recipient and would be required to report the subgrant to City B. However, City B does not have specific reporting obligations, nor does the contractor or subcontractor for the purposes of reporting on the Recovery.gov website."[3]

This in effect hobbles the accountability commitment of the Administration. The trail stops very high in the funding chain, making the ultimate destination of funds a mystery.

It is not sufficient to know HUD made a grant to New York, which then made a grant to New York City. We should know to whom the grant was ultimately made. This level of detail allows citizens with local knowledge to uncover if funds are being used in accordance with the law, revealing fraud and misuse.

Every dollar in the funding chain should be accounted for. Further, the reporting requirements do not stipulate how data should be provided. There is no guarantee that the complete dataset of recipient reports will be provided, or that they will be provided in a useful format.[4]

b) *Standardization*

It is currently unclear how data will be provided. To continue the spreadsheet metaphor, we don't know what the columns and rows look like. The Act requires initial recipients to base their reporting on the Federal Funding and Transparency Act. Thus, we expect reports will contain, award grantee names, amounts, program source, description, city, and state. But we do not know what data elements will actually be published, or the format in which we can expect it.

It would be helpful to know what Recovery.gov intends to provide, and in what form. That way, software developers can begin work on applications. Ideally data should be in XML format.

c) *Aggregation*

When information is standardized (the what, who, when and where of data), then it is possible to aggregate it.

---

[3] Office of Management and Budget, Initial Implementation Guidance for the American Recovery and Reinvestment Act of 2009, Feb. 18, 2009, pages 14-15, available at http://www.whitehouse.gov/omb/asset/aspex?AssetI=703.

[4] ARRA § 1512 (c) requires stimulus fund recipients report to awarding agencies how they have spent funds, there is no requirement in the Act that reports be made available to the public.

d) *Centralized Access*
Funding will be widely distributed, thus information will come from many sources. For the information to be useful, it must be searchable from central locations. Recovery.gov should function as a web search engine that houses every single reporting dataset. That does not preclude individual agencies publishing spending data on their websites. An analogy is to think of reports as books, indexed in a card catalog. As long as we know where the book is housed, it is possible to find it. Recovery.gov doesn't have to have all of the datasets, just the key for finding them.

Conclusion

With the passage of ARRA, the Administration and Congress made a commitment to citizens that the government would ensure transparency and accountability for how stimulus funds were spent. That task is only possible with the involvement of citizens – interested, technically skilled, and knowledgeable parties from across the political spectrum who want to participate in and collaborate with their government.[5] That community needs to know what data will be provided, and how to build the tools needed to make this effort work.

The President's Memorandum on Transparency and Open Government, cites three themes: transparency, participation, and collaboration.[6] The memo makes two important points developed in this testimony:

"Knowledge is widely dispersed in society, and public officials benefit from having access to that dispersed knowledge."

"Executive departments and agencies should use innovative tools, methods, and systems to cooperate among themselves across all levels of Government, and with nonprofit organizations, businesses, and individuals in the private sector."

These are the operating principles upon which Recovery.gov should build its transparency and accountability effort. Linking citizens with detailed, structured, and standardized data will make it possible.

Thank you.

---

[5] The Coalition for an Accountable Recovery, at http://www.ombwatch.org/car
[6] President Barack Obama, Memorandum for the Heads of Executive Departments and Agencies on Transparency and Open Government, Jan. 21, 2009, *available at* http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernmetn/