

Testimony Of

Eric Gillespie
Chief Information Officer
Onvia, Inc.

Before The

Subcommittee on Investigations and Oversight
Committee on Science and Technology
United States House of Representatives

At Hearing

**Follow the Money Part II: Government and
Public Resources for Recovery Act Oversight**

May 5, 2009

Introduction and Overview

Chairman Miller, Ranking Member Broun, and Members of the Committee, thank you for inviting me to testify today with this distinguished panel about oversight of the American Recovery and Reinvestment of 2009 (ARRA). My name is Eric Gillespie and I am the Chief Information Officer at Onvia, a 12-year-old, NASDAQ-traded information services company based in Seattle, Washington.

Simply stated, our business model provides a comprehensive view of government spending to companies around the country. We were purposefully designed for Local and Education marketplace and eProcurement.

Our business model is designed to facilitate procurement and eProcurement portals through an eProcurement portal for over 50,000 business users, and we drive procurement as an aside, this part of our business has been in place many years ago to help disadvantage.

On the other hand, we assist companies that do business with 9,000 State, Local and Education organizations such as also partner with organizations such as to track visibility for their members.

Our products track transactions from millions of goods and services transactions from across every industry vertical – from Construction, Engineering, and Architecture to Healthcare, Energy, Water, and Information Technology.

As draft versions of the stimulus bill were being published by the House and Senate at the beginning of the year, my team and I recognized that Recovery Act funds would be primarily distributed through existing programs, from Federal agencies to States, Counties and Cities, and

The Transparency Barrier

Transparency and reporting to both Congress and the public at large, with full disclosure of all entities, public and private, receiving funding from ARRA, tracking grants and subgrants, contracts and subcontracts, obligations and certifications, and authenticating the sources of this data, serves as the backdrop for defining the key challenges. Establishing accurate and timely job creation metrics adds yet another level of complexity.

The market is highly fragmented: there are more than 89,000 State, Local and Education entities across the country that are eligible to receive ARRA. In addition, the private sector has a role in procuring goods and services, and businesses that are creating jobs. These numbers are placed on an aggressive timeline.

The transparency barrier is a significant challenge for State and Local government as they manage the flow of funds at which these funds are being applied.

Consider the challenges of tracking subgrants and subcontracts in a highly fragmented market.



Congress, all attempting to transparency barrier Management and money that is spe funds – the contra Counties and Cit

American people are el; this ne Office of largely ignore e recipients of the g at a local level, used funds.

There are through the opacity. First, untangling the vast amount of unstructured data across these sundry entities is a Herculean task. Each entity has its own set of rules and workflow that address procurement, formats for solicitations, reporting requirements for contract awards, vendor qualification, vendor lists, and data persistence among many other areas. The categorization and compliance issues faced by the Federal government in implementing FBO.gov and the Federal Procurement

Data System (FPDS) over a number of years are similar, although given the wide dispersion of the State, Local and Education market the complexity is several orders of magnitude greater when attempting to track ARRA funds.

Once a canonical view of the entities is established, the second primary challenge can begin to be addressed: authentication of entities and their executives tasked with certifying the ARRA spending. A structured registration and validation process is required to ensure data integrity, along with a basic support center to handle basic compliance questions and simple account issues like stimulus funds with entities mixing funded by stimulus certain to be fraught with inaccuracies and compliance reporting is necessary.

The third challenge is the use of technologies, principally in the form of the expectations of the American people through relatively simple search engines are presumed in the context of information in the government market. The web from what has been described.

The fourth challenge is that the funds are being dispersed. An “aggressive” accountability goals of the Administration are being met, to create jobs as rapidly as possible, presents an opportunity for unprecedented waste and fraud. The Chairman of the Recovery Accountability and Transparency Board, Earl Devaney, estimated that \$55 billion of taxpayer dollars may be lost to fraud, which is particularly true at the State and Local level where the Administration has very limited visibility. That’s five times the entire GDP of Afghanistan; about the same as the GDP of Vietnam, Luxembourg, and Ecuador; and half the GDP of New Zealand, Egypt and Iraq.

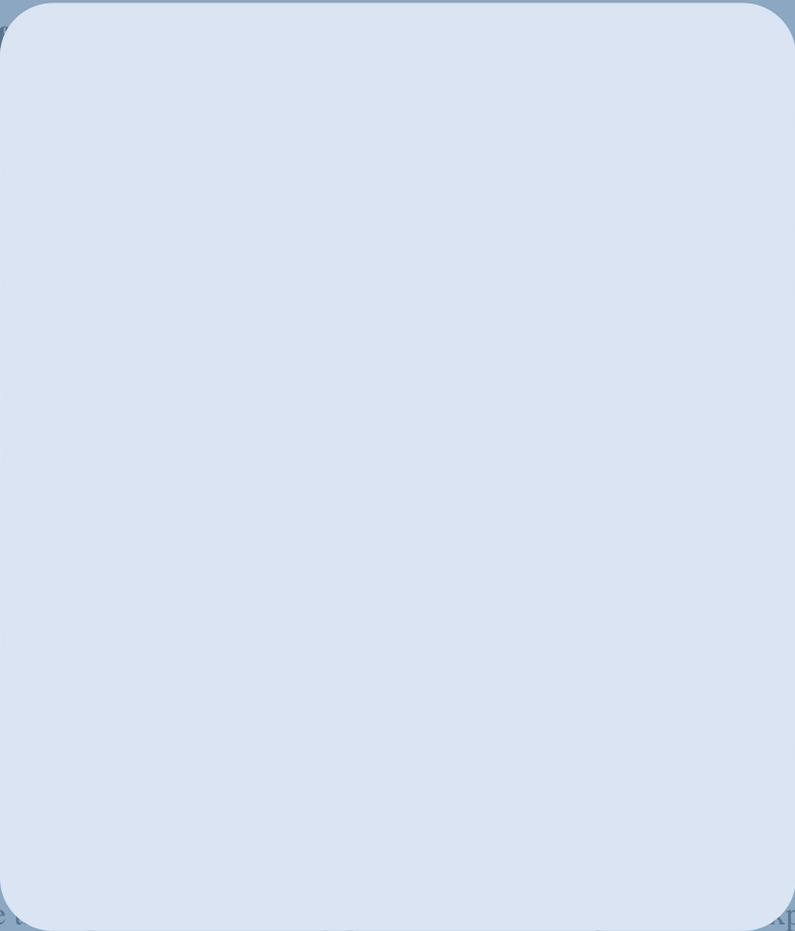
According to figures from the World Bank and the International Monetary Fund, the fraud alone in our ARRA spending will be the 60th largest economy in the world. Eliminating the transparency barrier with only a 1% improvement in fraud would save the American taxpayers \$550 million. With the aggressive timeline there is simply not an opportunity to spend months, or compared to some implementation cycles years, developing a delivery platform, and there is little opportunity to stop waste *before* it starts rather than relying on investigations *after* spending occurs.

Technology Platform

We appreciate the Recovery Act spending began solving it over hundreds of servers scanning and optimizing Freedom of Information many years of experience to meet these tracking these solutions.

When looking at spending on a Federal budget. Can be in granular detail; dependents, profit mortgage, and the

the purchase of goods and services occurs, the government marketplace is by far the largest “industry” vertical; citizens, businesses, non-profit organizations, State and Local agencies, and schools are involved in spending. However, the level of transparency and tracking on the expense side pales in comparison to that on the revenue side. From a technology perspective the IRS is able to sift through massive amounts of data on the revenue side because they have



ing to track ologically. We platform with echnologies, on-demand researchers have innovative ways record with

ired to track “side” of the flows of capital mber of aid on their expense side where

can drill down to their community and see the spending that is most important to them via zoom controls.

Given the time constrained “use it or lose it” provisions contained in the legislation the bulk of the projects we are currently tracking at *Recovery.org* are related to infrastructure and transportation.

With *Recovery.org* we have done precisely what the Administration has been encouraging with respect to the use of data. We took a large, unstructured dataset and turned it into something searchable and actionable.

The Complexity

The potential for data to inform decision-making is vast. However, the lack of standard formats and protocols for data collection and reporting across agencies and jurisdictions has made it difficult to aggregate and analyze data from multiple sources. An effort in recent years to create a common data format for the federal government has been unsuccessful. While the Coburn-Obama Act requires agencies to provide data in a quick, accurate and accessible format, the lack of a common standard makes this a difficult task.

The most significant barrier to data collection and analysis is the lack of a common data format. The Federal government has made significant investments in data collection and analysis, but the lack of a common standard makes it difficult to aggregate and analyze data from multiple sources. This is particularly true for State and Local governments, which often have their own data collection and reporting systems. The result is a fragmented data landscape that makes it difficult to get a complete picture of government spending and activity.

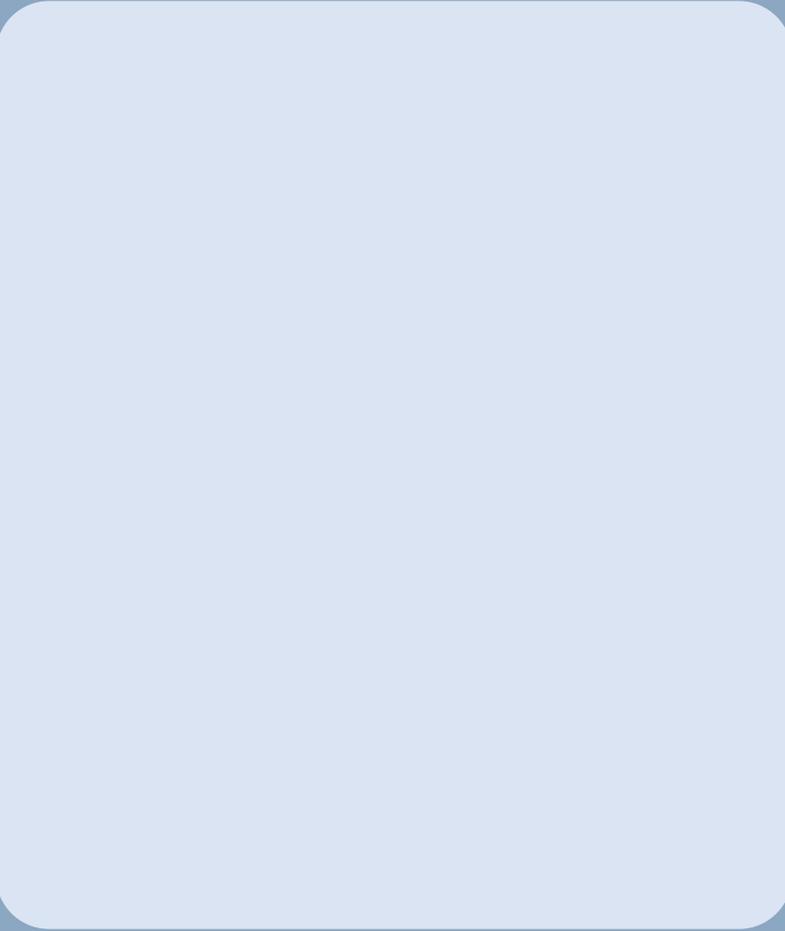
To further complicate requirements for timely and standardized data collection from State and Local entities, many of these agencies perform a combination of essential services ranging from public safety to maintenance of physical infrastructure. Unlike the Federal government's established taxonomies for tracking spending on goods and services, there is no universally accepted standard across these highly fragmented State and Local governments. What initially

appears to be a relatively simple set of functional spending categories is not easily mapped to a common State-by-State, Municipality-by-Municipality view.

For the moment let's assume the data aggregation and standardization problems can be solved for Recovery fund tracking, or for that matter any other dataset such as TARP fund tracking. Consider the mashup possibilities with census data, campaign contribution results, crime statistics, or tax information, to name a few views, and imagine the level of citizen engagement that might be generated.

Unleashing
about preventing
around those laud
well, the Recover
infusing capital fo
have the ancillary
efficient channels

There are
position on prome
Human Genome I
good. In addition
the GPS industry
community, and t
unleashed by the
sets and giving ac
for transparency.



ibility, nor solely
us far centered
ity. If executed
nomy by
ata it can also
vs to more

taken a leading
ney have used the
o the greater
amples include
to the geospatial
rted by data
ng large data
l gold standard

Without getting into deep technical details, suffice it to say that in order to maximize use and adoption the data must to be made available in formats with low barriers to use. There are excellent, free, non-proprietary formats such as XML, JSON, and YAML. There are other

standards such as XBRL which the Securities and Exchange Commission has adopted for electronically collecting financial results from companies.

Search engines, user interfaces, and infrastructure are all key to a successful technology-based product like *Recovery.gov* but the primary reason we were able to build *Recovery.org* in two weeks was our underlying, standardized taxonomy and data. Without the data none of the other technology would have mattered.

Recommendations for D

Representatives receive hundreds of millions of dollars including those of the federal government emphasized by the *Technology Solutions* program. These incredible volumes of government via technology innovation with the

It is difficult to spend the funds because very few Transparency Boards can perfect the process. I expect it to take the

I have several recommendations on the site:

receives measure This is further *on Information*. These engaging with their culture of web ism for stimulus untability and than 10 years to hile I wouldn't and transparency

- (1) A project impacting so many aspects of the government is not undertaken without some level of risk. Data will never be perfect. Accept that and get on with it. Adopt an implementation framework that is designed for nimbleness, such as Agile Scrum, to facilitate speed.

(2) A good user experience is paramount to success. Consider the many needs of the audience and distill them into a few basic, representative personas around which the site can be designed. View transparency and accountability as a leading “brand” with constituent touch points, of which *Recovery.gov* is perhaps the most significant in terms of its appeal.

(3) Define a simple common vocabulary for constituents, developers, Congress, the Administration, States and Cities. This should include terms like *appropriation*, *obligation*, *approval*, *certification* and *award*, among others. Establish a standard corpus of entities and information architecture to support simple data

(4) Choose a research as the format is chosen

(5) An initial begin process the broad interhone the appri

(6) Implement geography, Federal access project details of Recovery. With this, government as a result tax efficient.

(7) Operate a Recovery Act program assistance center to assist Federal, State and Local agencies with compliance, and assist potential recipients of funds by answering questions about securing grants or contracts. Learnings from this center should be used to inform further development of the site.

(8) As the data set grows, provide raw data feeds via APIs to the public at-large. It is not incumbent on the Federal government to create unique and interesting views of the data and, instead, by providing data to developers the public at-large can create engaging user experiences with the underlying data.

(9) Look for best practices and pockets of innovation across the public and private sectors which can be adopted. To be successful this can't be onerous on either State, Local and Education entities or the private sector contractors and subcontractors. It is also important to not create an i

The benefits of... tools to interpret
it, and the condui... pparency and
accountability. It... l have live,
searchable data in... received at
Recovery.gov wil

The communi... s an apt
description of wh

Conclusion

With that, ... an impossible
task, there is an e... a new era of
transparency, acc... to come in terms
of engaging in a c... ificant culture
shift for many wh... rounced at the
State, Local and Edu...

The technology is available to turn Recovery.gov into the flagship for government transparency and accountability. We fully support the important goals that Congress and the Administration have outlined, and we will continue to serve in any way we can.

Thank you again for inviting me to testify here today, and I look forward to answering any questions that you might have.

