

On the Road to Accountability and Transparency

The Building of FederalReporting.gov and Recovery.gov

Background

By any normal measure, the project should have failed. Indeed, the idea of developing two large-scale websites in a fraction of the usual time was enough to scare off 95 percent of eligible contractors from even bidding. An independent evaluation of the plan deemed it “extremely high risk” – fraught with numerous challenges, any one of which could torpedo the entire venture. Just coordinating input from 28 federal agencies would be a significant achievement in governmental operations.

Washington, of course, has seen its share of failed projects and programs, piled one on top of the other. But in this case, Congress and the White House decreed that neither failure nor delay was an option. The message: Get the job done!

Against that backdrop, the technical staff of the Recovery Accountability and Transparency Board and outside contractors forged a close partnership to build version 2.0 of Recovery.gov and FederalReporting.gov. Those systems are the twin pillars of the effort to fulfill the unprecedented levels of accountability and transparency mandated by the American Recovery and Reinvestment Act of 2009. In a nutshell, determination, commitment, improvisation, careful planning and a lot of elbow grease accomplished the daunting task.

But to comprehend the magnitude of the accomplishment, a look back is essential. In perhaps its most important decision, the Board - established by the Recovery Act to oversee reporting and spending - decided early on to cut through bureaucratic red tape and to deploy innovative management techniques. An agile IT team was assembled from other agencies, contracting was expedited, and a core management team was created to assess risks and to consult with other key government players.

Working with the IT team, the Board established a flexible but disciplined strategy with specific milestones and a so-called “go-live” date for the two websites. The IT team deployed critical path and adaptive management techniques, enabling it to stay on track but to adapt as necessary. These techniques were critical given the extreme time deadlines. Technology was kept simple, web-based and scalable so that the websites could handle any increased demands.

The job at hand was critical to the nation. By signing the Recovery Act into law in February 2009, President Barack Obama set in motion the largest federal economic plan since the Great Depression. At the same time, the law created a new era of accountability and transparency,

requiring recipients of Recovery funds to submit detailed data reports on their awards following each calendar quarter, with the data publicly released 20 days later.

The Office of Management and Budget (OMB) mandated reporting to begin on October 1, 2009, which presented the most daunting challenge of all to the Board. Within the federal government, there was no data collection or reporting system that could handle the massive volume of information that would be coming in. What is more, the General Services Administration (GSA) informed the Board that a solicitation and award of this magnitude normally takes up to nine months.

But the Recovery Board had only five months to accomplish everything - contracting, development, construction, testing and deployment of both websites - all while Congress, the public, the media and countless interest groups were watching and waiting. Meanwhile, the Board also had to manage version 1.0 of Recovery.gov, the website established early on by GSA and the Office of Management and Budget (OMB) to provide Recovery information to the public. The Board, consisting of the Chairman and 12 Inspectors General, insisted that all information must be presented transparently without partisan overtones.

The Board moved its operations and staff into permanent offices in April 2009. The next step was critical: the Board retained the MITRE Corp., a prominent national research organization, to do an independent review of version 1.0 of Recovery.gov. MITRE researchers reported that a new site with much more functionality was needed, along with a separate new website for collecting the mass of data required by the Recovery Act. Equally important, MITRE cited a lack of experienced personnel necessary for managing the development of two new sophisticated systems. The report also said that the normal federal acquisitions process risked delaying the project beyond the hard deadlines.

“Overall, we believe this effort to be extremely high risk,” the review concluded. “The current state of the management and acquisition models, along with the high expectations and short timeframe in which to deliver, are the fundamental constraints which create this high risk environment.”

But how should the recipient data and information be displayed? What website features would best serve the public? To answer these and many other questions, the Recovery Board teamed with OMB and the National Academy of Public Administration (NAPA) to hold a week-long electronic town hall meeting in the spring of 2009. The goal was to draw on ideas from the public and IT professionals for remaking Recovery.gov.

Participants in the online dialogue included prominent names in the world of Internet innovation, perhaps none more notable than Tim Berners-Lee, the very creator of the World Wide Web. Erik Wilde, a leading professor at the prestigious University of California Berkeley School of

Information, and digital media guru Tim O'Reilly, who coined the term "Web 2.0," also contributed to the discussion.

The forum yielded more than 500 ideas, most of which revolved around the following:

- Huge amounts of data should be the site's main content.
- The site had to be dynamic, highly interactive, and visually pleasing.
- Features should include sophisticated search capabilities and mapping technology.

The Board also arranged extensive focus groups to continue taking the public's pulse and met with a range of stakeholders to determine their interests and needs. A vision of a website that would contain rich content with a focus on local stories of Recovery spending rather than the Washington establishment began to emerge. But again, due to time constraints, the Board would effectively require contractors to propose multiple designs and ideas as the focus groups and stakeholder meetings were happening, and then adapt those plans as needed.

By the time the final MITRE report was delivered in May 2009, the Board was already addressing many of the challenges identified. The Board, for instance, had begun to put together a team with the knowledge, experience and skills to manage the considerable risks of the project. Moreover, the Chairman of the Board met with the head of GSA to emphasize the necessity of expediting the acquisition and contracting process. GSA responded by assigning the agency's chief procurement officer - its top contracting official - to ensure the Board's contracting requirements were met legally, in full, and on time. GSA had never established such an arrangement before.

The Board then created a tightly focused project management environment that could be as efficient and streamlined as necessary to meet the extreme deadlines for standing up two complex Internet sites. Conflicts had to be kept to a minimum—and they were.

Under normal circumstances, such a project would proceed in sequence: Specific requirements would be determined according to the needs of both the client and prospective users, a prototype would be developed, then tested, then readjusted, tested again, and so forth until a final product meeting client approval is delivered.

However, the time constraints permitted no such luxury. Things had to happen simultaneously. As one contractor later noted, "Basically, we were building the ship as it was leaving the port."

FederalReporting.gov

Given the hard deadlines, developing and testing new technologies specific to the project would have been impractical. Therefore, the Board and its IT team decided to adapt existing technologies in new ways. Initially, they thought of joining together existing legacy computer systems within federal agencies to collect and house data from recipients of Recovery funds. But that course was ultimately rejected because of data, timing and linkage issues that could lead to project failure.

Instead, the Board and its IT managers decided to develop the data collection, or inbound reporting site, by buying into an existing, competitively bid contract that the Environmental Protection Agency uses for its own web-based reporting system. Built by CGI Federal of Fairfax, VA, the system appeared capable of being upgraded and expanded to meet the technical specifications of recipient reporting.

In addition, the Recovery Act required that a wide field of data points and other information be collected. Among the numerous requirements:

- Relevant economic, financial, grant and contract details;
- Information about the competitiveness of the contracting process;
- Information about the process used for award of contracts;
- A summary of any contract over \$500,000.

Obviously, the Board expected an extremely large number of recipient reports would be submitted to the inbound site. But no one knew precisely what that number would be. OMB initially thought that between 40,000 and 250,000 recipient reports would be filed. With such uncertainty, the Board decided to build for the high end of the estimate.

As it turned out, the new inbound reporting site would resemble no existing governmental system. Instead of having each federal agency that awarded Recovery contracts, grants and loans collect data from their recipients, the Recovery Board and OMB opted to create a single, centralized reporting system to be called FederalReporting.gov. All recipients would file data into this one repository, which would be accessible to federal agencies so that they could review submitted data, as required by the Recovery Act.

FederalReporting.gov also would be the first to handle government contracts, grants and loans all in one system – convenient for data trackers, but a potential nightmare for IT system builders. For instance, federal contracts and grants each involve different user-communities as well as rules and policies. This would entail different coding, prototypes, testing, and approvals regarding data for each type of award.

The system had to be extremely flexible. The Recovery Act says recipients are required to file their reports in the first 10 days following the close of a calendar quarter. But most businesses don't gather all their financial data in the initial days after a quarter's end. This meant FederalReporting.gov would have to be able to absorb a likely surge of submissions in the latter part of the 10-day window. However, this flexibility did not extend to changes in the scope of the project. Only one major change, allowing for bulk reporting by states, was added during the development cycle.

Moreover, the system had to be secure to prevent hackers from breaking in as well as non-recipients from entering the system and filing erroneous or false data. But security had to be elastic as well as ironclad. After the 10-day window closed, recipients still had to be able to review and correct their respective data in the system, but at a certain point recipients had to be locked out while federal agencies performed independent data reviews. To accommodate this level of security, new business logic would have to be written.

Under the Board's timetable, FederalReporting.gov had to be up by the third week of August so that Recovery award recipients could register, as required, in advance of the first reporting period beginning October 1, 2009.

Finally, unlike other governmental reporting systems, FederalReporting.gov had to be able to accept data from not just states and industry but also from mid-sized corporations, universities and even mom-and-pop-type businesses. The system, therefore, had to be understandable to a wide array of users and stakeholders with differing levels of computer and online skills.

Taking direction from the Board's technology team, CGI Federal had only six weeks to build and deploy FederalReporting.gov. Moreover, CGI Federal had to establish and train a Help Desk staff and develop and distribute training materials for federal agencies and recipients that would be using the site.

Working from the EPA reporting system it had built, CGI Federal was able to leverage not only an established technical platform, but also the best practices and methodologies the contractor has developed on other projects worldwide. The solution would be hosted as a "cloud," or virtualized service, in CGI Federal's Tier III certified data center.

The in-bound software development team used a methodology known as Rapid Application Development (RAD) techniques to quickly create and implement prototype application components while other design efforts were under way. The registration component of FederalReporting.gov, for example, was put in place while design of the reporting functionality was still going on. With input coming daily from the Board's IT team, CGI Federal also used a team management process known as Joint Application Design (JAD) to design requirements on the fly. In some cases, requirements defined in morning JAD workshops and built that evening

were then reversed the next day when conflicting requirements arose. This kind of adaptability allowed for rapid responses to evolving requirements.

Recovery.gov

For building FederalReporting.gov, the Board directed and worked closely with CGI Federal to significantly expand the kind of data reporting system the company had built for EPA. But finding the right contractor for Recovery.gov 2.0 wasn't so obvious or simple. Coordinating with GSA, the Board decided to use a procurement vehicle known as the Alliant Government-Wide Acquisition Contract, or GWAC, a prequalified group of seasoned and technologically capable government contractors. The Board's request for proposals had stated, "Recovery.gov 2.0 Website Redesign," but the task was really much more than that. The new site required, among other elements:

- A secure, highly available, fault-tolerant data center infrastructure;
- An ability to assimilate disparate data sets from more than a dozen different sources;
- Features that allow quick data analysis and dynamic reporting.

All of these elements would have to be built from the ground up.

But because of the size of the project and extremely tight deadlines, 56 of the 59 Alliant members didn't even bid, including one well-known computing giant. Working with GSA, the Recovery Board evaluated the three submitted proposals on four factors: management approach; technical expertise; past performance; and price.

In July 2009, Smartronix, Inc., a Maryland-based firm experienced in federal and military IT projects, won the competition. Smartronix's proposal included three subcontractors – Synteractive, TMP Government and KPMG. Team Smartronix's proposal meshed a robust technical solution with a critical understanding of tools needed to enable citizen-government collaboration. Version 2.0 of Recovery.gov had to go live before the end of September.

As it had in developing FederalReporting.gov, the Board put in place a management process that allowed for quick decisions, adaptability, and a can-do attitude among those involved in developing Recovery.gov 2.0. The challenges were many: technological complexity, evolving stakeholder requirements, evolving data changes and data requirements, and the need for collaboration across multiple government agencies and federal contractors.

Due to the velocity of the development process, each day's application developments were tested at night to identify performance issues. The work schedule for the team included weekends and holidays. The importance of this project to the nation produced an unusually high level of motivation and personal commitment from everyone involved.

Smartronix quickly implemented a centralized Program Management Office (PMO) that oversaw the daily operations of several working groups known as Integrated Project Teams (IPTs) and Orchestration Teams. The IPTs were responsible for application, data, infrastructure, and creative design, while the Orchestration Teams were responsible for systems engineering, quality management, security, and strategy/communications. IPT leads had the autonomy to make all decisions related to their specific team, and the Orchestration Teams ensured continual communication across all IPTs. The PMO established a centralized portal for all of the teams, the Recovery Board, and CGI's Inbound Reporting Team. This portal ensured a close communication of priorities and project status to all stakeholders.

Infrastructure

Team Smartronix's Infrastructure IPT designed, ordered, installed, cabled, configured, patched and secured more than 30 enterprise server, storage, and networking devices. The team was also responsible for everything from the external public router to the underlying storage network. Team members worked extended shifts to get the system in place to meet the aggressive deadlines.

Typical solutions of this nature can take up to four or five weeks just in the logistics of the procurement and delivery phase. Team Smartronix worked closely with its delivery partners in ensuring urgent priorities were met, thus enabling the data center solution to be built in less than five weeks. Although rapidly assembled, every interconnection was meticulously documented and labeled as infrastructure engineers performed scalability, load balancing and fault-tolerance testing. The team also conducted peer reviews with all vendors to ensure the solution would operate at the highest levels of efficiency. This foundation later contributed to the high availability of the website.

Application

This solution would result in one of the world's largest Microsoft Office SharePoint Server (MOSS) for Internet deployments. One core requirement for Recovery.gov was a highly tailored Content Management System (CMS) that met unique publishing and workflow requirements. The Recovery Board needed to control website content without requiring developer support for each release.

Among the critical first steps was the Board's decision to use Microsoft SharePoint as the collaborative platform for Recovery.gov 2.0. SharePoint, a multi-faceted collection of products and software elements that can be used to host websites, comes out-of-the-box ready for integration with other technologies and capabilities. It is also extremely scalable and can serve as a CMS and, equally important, can offer in conjunction with other Microsoft products a robust search function.

To meet the deadlines, the Application IPT needed an infrastructure immediately so it could commence development and testing. By utilizing Amazon's Elastic Computer Cloud (EC2), the team was able to have a 20-server deployment in place in a matter of hours, while the actual infrastructure was still being built. This was the first time that the EC2 solution was used with the Microsoft Platform in federal space. Team Smartronix's subcontractor, Synteractive, led the efforts.

The Cloud was used exclusively for the test and development environments. This environment enabled the team to have a representative logical/virtual architecture that matched the physical architecture being built in parallel. When the physical architecture was completed, it only took a few days to take the virtual architecture and replicate it to the physical. By synchronizing the test and development environments with the production environment, the application team was able to quickly adjust to the rapidly changing requirements.

Creative Design

The Recovery Board periodically met with experts for guidance on how to develop Recovery.gov. These experts provided the Board with their valuable time in the interest of helping their country. They included, but were not limited to, Ben Shneiderman, founding director of the Human Computer Interaction Laboratory at the University of Maryland; Jonathan Lazar, a computer science professor and director of Towson University's Universal Usability Laboratory; and Edward Tufte, professor emeritus at Yale University, where he taught courses in statistical evidence, information design and interface design.

Shneiderman urged Board Chairman Earl E. Devaney to work closely with the Board's stakeholders—the public, Congress, the media, and watchdog groups. Lazar said it was critical for the Board to ensure that Recovery.gov met the needs of the disabled and complied with Section 508 requirements of the Rehabilitation Act. Tufte, meanwhile, said that the Board should not be overwhelmed by “feature-itis.” Content, he said, should be the Number One priority, noting that data from recipients would amount to “a content bonanza.”

The Creative Design Team, led by TMP Government, had to operate under intense time pressure since the design was a critical milestone delivery significantly impacting all IPT teams. This was amidst constantly evolving requirements from the many Recovery stakeholders. The design team had a base set of requirements from the contract solicitation, but these were insufficient to meet the desires of all stakeholders.

Working closely with the Strategy IPT and the Recovery Board, the Creative Design Team developed a set of hypothetical user segmentations and derived the initial requirements from this. These user segmentations included the press, state and local governments, federal agencies, Congress, the Recovery Board, watchdog groups, transparency developers, and the general public.

Syneractive developed the user-segmentation approach based on questions such as:

- How would the average citizen want to look at the data?
- What information would the public want to retrieve from Recovery.gov?
- How would this differ from what the press might want, or special interest groups, or Congress?

Next, the Creative Design Team began building templates and then began a fact-based approach including focus groups, stakeholder meetings, usability testing, expert panel design reviews and 508 compliance testing. Focus groups were set up in multiple cities across the United States and represented several citizen demographics.

But before results even came in, the Creative Design Team was developing multiple design templates for each specific part of the site. The various options served as available back-ups in case the first selected template proved unsuitable. As the results poured in from all the fact-based analyses, the team had to rapidly adjust its designs to meet the evolving requirements.

Syneractive collaborated with lead designer TMP to ensure visual design complemented the strengths of Microsoft SharePoint. Meanwhile, TMP began developing a highly robust and flexible design that focused on usability, including sixteen 508 compliance requirements and the deployment of digital techniques to ensure that the complex data and information was easy to find and understand through simplified visualizations. TMP designed a graphical user interface for a mapping service that allows users to see exactly where Recovery money is being spent, from a national level all the way down to ZIP codes.

Because of the complexity and depth of information that would be provided to visitors of Recovery.gov, TMP also designed a series of online tutorials for easier understanding of how data is collected, what the data means, and how to best use Recovery.gov.

Data Intake and Display

Typical large-scale data warehouse efforts require months of upfront design and have clearly defined reporting requirements. These efforts can take up to a year or more to implement, especially if they involve data sets with varying reporting frequencies and evolving changes to the underlying model. Additional hurdles involved ensuring that the functionality would scale to millions of users, and would integrate with the Content Management System and the geospatial/mapping framework planned for Recovery.gov.

The data sets driving Recovery.gov come from multiple sources and a variety of formats. Some data, such as the recipient reports from FederalReporting.gov, come from a relational database that is replicated directly into a private enclave at Recovery.gov. Other data sets come in on a

daily, weekly, or monthly basis in a variety of formats including Excel Spreadsheets, XML data loads, and web service feeds. Each data set has its unique challenges for extracting, loading and linking the data into the data warehouse.

In all, more than a dozen different data sets were integrated into the Recovery.gov solution and more are being added every month. The initial launch ten weeks into the project included eight distinct data sources. On top of the challenge to provide a public-facing website, the team had to build an analyst-facing reporting solution. This would enable the Recovery Board's data team to perform *ad hoc* reporting and analysis to support the needs of the administration, Congress, the media, and agency stakeholders. These reports had to provide business intelligence on data quality, data anomalies, and financial analysis. The overall solution utilizes MOSS 2007, ESRI Geospatial Information Services, ESP FAST Search, Business Objects, SQL Server 2008, Systems Center Operations Manager, Forefront Antivirus, JackBe MashUp Server, firewalls, VPNs, IDS/IPS, Cisco switches, Fiber Channel SANs, DELL Server hardware, F5 BigIP load balancers, Amazon EC2 Cloud and EBS storage, Google Services, Bing Services, and custom-developed web components.

Geospatial/Mapping

Results from focus groups and stakeholder meetings made one thing abundantly clear: People wanted to know how Recovery money was impacting their own neighborhoods. Responding to this insight, the Board's technical staff and Team Smartronix worked to ensure that the site provided a comprehensive geospatial, or mapping, capability focused on the local impact of spending.

A geographic information system (GIS) would enable people to see, understand, evaluate, and give feedback on government decisions. GIS mapping would show where Recovery spending was actually taking place and enable evaluation of intended outcomes of investments in local communities.

Integrating its GIS software into the data display solution that Team Smartronix was building, California-based ESRI gave government leaders and citizens in a matter of months the ability to create interactive maps that displayed the status and cost of Recovery projects throughout the nation.

With GIS, the depth of information available for viewing goes well beyond simple consumer mapping applications that may only display limited geospatial attributes. The sophistication of GIS is in its ability to access and analyze data from a multitude of geo-databases and make that information accessible in an intuitive way. ESRI's president says it this way: "The maps are generated using ESRI's geographic databases, which are populated by reports from recipients. ESRI systems can read almost anything that a person or machine feeds into the software and then turn it into a digital map."

The great significance of GIS is that it incorporates spatial analysis tools that assist in determining the cost of recipient projects and highlighting fund distribution in given areas, including states, counties, congressional districts and zip codes. GIS also serves as a performance measurement tool for evaluating expenditures and keeping the public informed.

In the end, the mapping solution proved to be one of Recovery.gov's most popular features.

Security

Team Smartronix's Security IPT was responsible for boundary protection, application and server vulnerability assessment, risk mitigation, intrusion detection and prevention services, log management, systems center management, operating system security implementation guidelines, virtual private network segmentation, identity management and user provisioning, patch management, and enterprise antivirus configuration.

The solutions and processes had to be planned and designed in parallel with the application and Infrastructure IPTs. The result was a security solution delivered in six weeks, which received full authority to operate from the Board in nine weeks. Such a quick turnaround on security measures for a public-facing website was extraordinary.

Social Media and Other Web 2.0 Features

TMP also designed and branded four social networking sites for Recovery.gov - YouTube, Facebook, MySpace and Twitter. Today, Recovery.gov has thousands of Facebook and Twitter followers and a dedicated YouTube channel for streaming video blogs and training. Few federal websites have this level of social media capabilities. Recovery.gov enables people to easily share content via more than 100 social commentary sites such as Digg, Delicious, and Yammer. The site utilizes Cloud Software as a Service (SaaS) capabilities such as Google Translate, Google Analytics, Bing custom search, and ESRI's software, ArcGIS Online.

Google Translate allows the site to be translated in more than 50 languages. Google Analytics provides the usage and audience segmentation analysis. The site uses a customized Bing Scope that searches thousands of recipients' websites and returns job opportunities. ESRI's ArcGIS Online enabled cloud-based geospatial layers for street level and aerial views, significantly reducing the geospatial hardware footprint. The site enables citizens to submit feedback or fraud, waste and abuse claims directly. It introduces polling as a mechanism to help the Recovery team improve site quality and audience analysis.

Standing Up to the Test

For the Recovery Board's technical staff and contractors, the first true test came on August 17, 2009, when CGI Federal deployed FederalReporting.gov so that Recovery award recipients around the country could begin registering for data reporting. Overall, registration went quite smoothly, largely because OMB, the Board and CGI Federal educated recipients about the system and its functions.

But a bigger test lay ahead, on October 1, when recipients began filing what turned out to be more than 130,000 separate data reports. Throughout the reporting cycle, FederalReporting.gov experienced no software defects that prevented either recipient reporting or agency reviews. Some recipients, however, failed to submit timely reports because they simply did not understand the new processes and policies established for the reporting cycle. Those problems would largely disappear in the next reporting period as recipients became more comfortable with FederalReporting.gov.

Recovery.gov version 2.0 was launched on September 28, 2009. On October 30, as required by the Recovery Act, the site posted and displayed an enormous amount of data transferred smoothly from FederalReporting.gov. Many website experts applauded the savvy look and increased functionality. Both new web-based reporting systems have continued to run smoothly for all reporting periods so far.

The Board's IT team members universally say they learned two valuable lessons from the strategy of developing components of both websites simultaneously instead of serially:

- Developing the usual monolithic, one-size-fits-all, single, customized technology to solve a problem isn't necessary. Quickly and imaginatively integrating existing technologies can work as well or even better.
- Deploying individual tools and components whenever they are ready and integrating others as they become available can work just as well as waiting until all pieces are ready to go.

A New Standard

The Recovery Act provides funds to more than 275 federal programs. In the past, these programs would have relied on their own separate data-reporting systems, but because of the creation of FederalReporting.gov and Recovery.gov, a single, two-part system is now used to capture and display data for all of these programs. That kind of innovation continues to this day.

In May 2010, for instance, the Board announced that Recovery.gov had moved to a cloud computing infrastructure, a technology that allows for more efficient computer operations, improved security and reduced costs. Recovery.gov became the first government-wide system to

move to the cloud, using Amazon's EC2 commercial cloud that provides infrastructure as a service.

Recovery.gov has now been honored for innovation at least six times by local and national organizations including the Ad Club of Metropolitan D.C. and the 14th annual Webby Awards. Simply put, Recovery.gov 2.0 has raised the standard for accountability and transparency in the federal government. As a recent Newsweek article on the re-launch of the site noted:

“The result is the current incarnation of Recovery.gov—which, as anyone who has spent significant amounts of time scouring government websites for information will tell you—is perhaps the clearest, richest interactive database ever produced by the American bureaucracy.”

###