Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Government Data Sharing Community of Practice

Conference Minutes

April 28, 2014

The MITRE Corporation, 7525 Colshire Drive, McLean, VA 22102

http://www.gao.gov/aac/gds_community_of_practice/overview

Background on the Community of Practice

In January 2013, GAO cohosted a forum with the Council of the Inspectors General on Integrity and Efficiency (CIGIE) and the Recovery Accountability and Transparency Board to explore using data analytics—which involve a variety of techniques to analyze and interpret data—to help identify fraud, waste, and abuse in government programs. Forum participants included representatives from federal, state, and local government agencies as well as the private sector. Through facilitated discussion, forum participants identified a variety of challenges that hinder their ability to share and use data. Among other challenges discussed, forum participants discussed the difficulty of balancing privacy laws and increasing the transparency of data. A summary of the key themes from the forum is published at http://www.gao.gov/products/GAO-13-680SP. To continue the dialogue on issues related to coordination and data sharing, GAO formed the Government Data Sharing Community of Practice.

Keynote

Gene Dodaro, Comptroller General of the United States, U.S. Government Accountability Office

Mr. Dodaro discussed the importance of data analytics in helping to correct the government's unsustainable fiscal path by curbing improper payments, identifying efficiencies, and helping smarter targeted investments, among other things. Mr. Dodaro also noted the positive developments in data analytics and discussed the Data Accountability and Transparency Act, agreeing with its goals to create data standards and solicit broad stakeholder input.

Mr. Dodaro discussed last year's Data Analytics Forum, highlighting the challenges and opportunities identified and the next steps agreed upon by cohosts, including the formation of a community of practice (CoP). He mentioned the two previous CoP sessions that have been held on Legislative and Technological Challenges in data sharing.

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Panel 1: Independence in the Age of Analytics

Moderator and Panelist
- Kathleen Tighe, Inspector General, U.S. Department of Education; Chair, Recovery Accountability and Transparency Board

Panelists
- Eric Holbrook, Assistant Director, Financial Management and Assurance, U.S. Government Accountability Office
- Lee H. Giesbrecht, Director, Statistical Operations Division, Office of Inspector General, U.S. Department of Veterans Affairs

Presentations

Kathleen Tighe - Inspector General, U.S. Department of Education; Chair, Recovery Accountability and Transparency Board

Ms. Tighe opened the panel with a discussion of challenges to independence posed by data analytics, data warehouses, and analytical tools. Specifically, Ms. Tighe described how the Department of Education Office of Inspector General (ED OIG) approached several instances where ED OIG's data analytics were of interest to an audited entity.

- For example, according to Ms. Tighe, ED OIG staff developed a data warehouse that an audited entity within the Department of Education, Risk Management Service, expressed interest in using. ED OIG staff weighed approaches for balancing independence and resource management and decided to pursue a Memorandum of Understanding (MOU) with Risk Management Service. The MOU stated that although the ED OIG and Risk Management Service shared a common data warehouse, they had separate roles and intended uses. For example, the ED OIG provided the hosting environment, which included security certifications, and Risk Management Service provided a SAS statistical application that allowed both entities to conduct data analysis.
- In another instance, Ms. Tighe said OIG staff developed automated tools that allowed OIG staff to search information on students receiving loans and schools participating in the financial aid program. ED OIG also developed a tool to be used to identify activities of fake students who sign up for college classes to receive disbursements of federal student aid money. Department of Education officials heard about that tool and expressed interest in using it. However, ED OIG staff had concerns about sharing the tool because ED OIG had developed it and, if the agency adopted it, ED OIG might subsequently have to audit it. Ultimately, ED OIG staff spoke with department officials about how the tool was developed, but did not share the tool itself.

Additionally, Ms. Tighe described challenges generated by the tools—namely, developing priorities to decide which potential wrongful activities to pursue and identifying the best approach to address potential wrongful activity when ED OIG staff have limited resources.

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

<u>Eric Holbrook - Assistant Director, Financial Management and Assurance, U.S. Government Accountability Office</u>

Mr. Holbrook opened his presentation by discussing how the Yellow Book (*Generally Accepted Government Auditing Standards*) evolved from a rules-based format to a conceptual framework approach. When developing previous iterations of the Yellow Book, GAO determined that the highest areas of risk would likely be certified public account (CPA) firms that conduct roughly 35,000 to 40,000 single audits per year. Instead of proscribing what auditors can or cannot do, the Yellow Book established a framework for auditors to use their professional judgment when assessing threats to independence. Mr. Holbrook said this approach seemed to be working well. Mr. Holbrook noted that the Yellow Book has better developed independence sections for CPA firms than for OIGs, and perhaps things could be added to the 2015 Yellow Book update to assist the OIG community.

<u>Lee H. Giesbrecht - Director, Statistical Operations Division, Office of Inspector General, U.S. Department of Veterans Affairs</u>

Mr. Giesbrecht provided examples of work done by his department's OIG and considerations regarding independence. Mr. Giesbrecht said the OIG is examining data for duplicate payment related to healthcare received by veterans from doctors who are not part of the Department of Veterans Affairs (VA). According to Mr. Giesbrecht, the Health and Human Service OIG is doing similar types of work using Medicare data, and the VA OIG is looking at ways to share the data with the Centers for Medicare & Medicaid Services OIG, though there are some questions about independence.

In another example, Mr. Giesbrecht discussed how OIG staff reviewed the statistical aspects of VA's Improper Payments Act report. The first year, VA did not do well; then sought feedback from the OIG; and the following year made improvements. The outcome in this case was positive; yet trying to determine how many data and analytic tools an OIG can share with the audited entity without affecting its independence remains an open question.

Selected Questions from the Moderator and Audience

<u>Question - To clarify, OIGs can share tools and outputs of analytic tools with programming units?</u>

Mr. Holbrook said there is nothing in the Yellow Book that would specifically prohibit sharing. However, Mr. Holbrook noted that facts and circumstances matter, and any potential sharing needs to be very carefully dealt with. Generally, under the conceptual framework of the Yellow Book, data sharing need not limit auditors if they can perform work without impairment to independence.

<u>Question - Could you discuss potential strategies for putting controls in place that would allow OIGs and agencies to share data, while protecting OIG independence?</u>

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Ms. Tighe noted that transparency is a good approach. Having a memorandum of understanding that clearly laid out roles and responsibilities was a way to protect her office in the example she discussed.

Question - An audience member noted that the member's organization develops data sets and is frequently asked by audited entities to use the data sets. What are some of the risks in this situation?

Ms. Tighe said the ED OIG sometimes reminds audited entities that it is using the entity's own data to conduct analysis—so the audited entity has access to the data. The more an auditor manages or analyzes the data, the harder it becomes to share.

The audience member noted that they sometimes provide data in paper form. Other times, they refer audited entities back to their own data.

Panel 2: Leveraging Data Analytic Resources and Measuring Effectiveness

Moderator
- Carrie A. Hug, Director of Accountability, Recovery Accountability and Transparency Board

Panelists
- Bryan Jones, Director of the Data Mining Group, Office of Inspector General, United States Postal Service
- Eric Schweikert, Director of Compliance Analytics Pilots, Internal Revenue Service
- Gordon Milbourn, Principal, The MITRE Corporation

Presentations

Bryan Jones - Director of the United States Postal Service Office of Inspector General Data Mining Group

Mr. Jones opened the panel presentations with a discussion of his role as the Director of the United States Postal Service's (USPS) OIG Data Mining Group and his experience with developing a data analytics program at USPS OIG. An advantage of his role at USPS is not being in one component group for any length of time, which allowed him to see broadly across the organization and not get mired down in "this is how we always do it."

Mr. Jones said the key to developing a successful program is to gain the buy-in of others. To do so, the first step is to show people the program's potential. Start with a problem that needs solving and approach the problem first from the perspective of the business question. Then determine the availability and accessibility of data required to respond to that need. It is also important to consider what data might be available from external sources, including other government and nongovernment sources. For example, in an effort to learn more about scan data, USPS partnered with companies, such as Netflix, Amazon, and Game Fly, to learn their processes and develop best practices to bring back to USPS.

Another key component of developing a successful analytics program, according to Mr. Jones, is to find early adopters on the frontline front line of the organization. Early adopters can provide feedback and help promote the analytical tool within the organization. Implementing a data analytics program requires a culture change, which can often create tensions. It is important to take a nondefensive posture and present the data in a way that promotes transparency and alleviates tension.

Ultimately, Mr. Jones concluded that the analytic tool must be made useful for auditors, investigators, and decision makers. The focus should not be on claiming results, but helping deliver a model that will help others claim results. Return on investment should be measured, not by the dollar effect, but by the number and frequency of people using the tool to better demonstrate to management and users what analytics can do for the organization.

Eric Schweikert - Director of Compliance Analytics Pilots, Internal Revenue Service

Mr. Schweikert discussed ways to measure outcomes of data analytics programs and how these outcomes can be organized to provide an overall view of identity theft. For example, measuring performance based on identifying more instances of identity theft can be a positive indicator in that more fraud is being identified; however, it could also mean that there is more identity theft being committed.

Mr. Schweikert emphasized the need to look at system performance more holistically. Where fraud is stopped in the system is important to the Internal Revenue Service (IRS), and it focuses on using analytics to identify fraud at the preacceptance and prefund stages because it is more difficult and costly to recover fraudulent refunds after disbursement. Data analytics are also helping to improve the balance between the need to stop a fraudulent payment prior to disbursement and the need to make proper payments in a timely fashion.

Additionally, Mr. Schweikert highlighted the difference between measuring outcomes for data analytic modeling and measuring system performance. Measuring system performance will entail including data points that are not identity theft. Outcomes based on data modeling require a high level of accuracy in detail in output. However, Mr. Schweikert emphasized the importance of considering the broader picture of how well the system performed in identifying noncompliance.

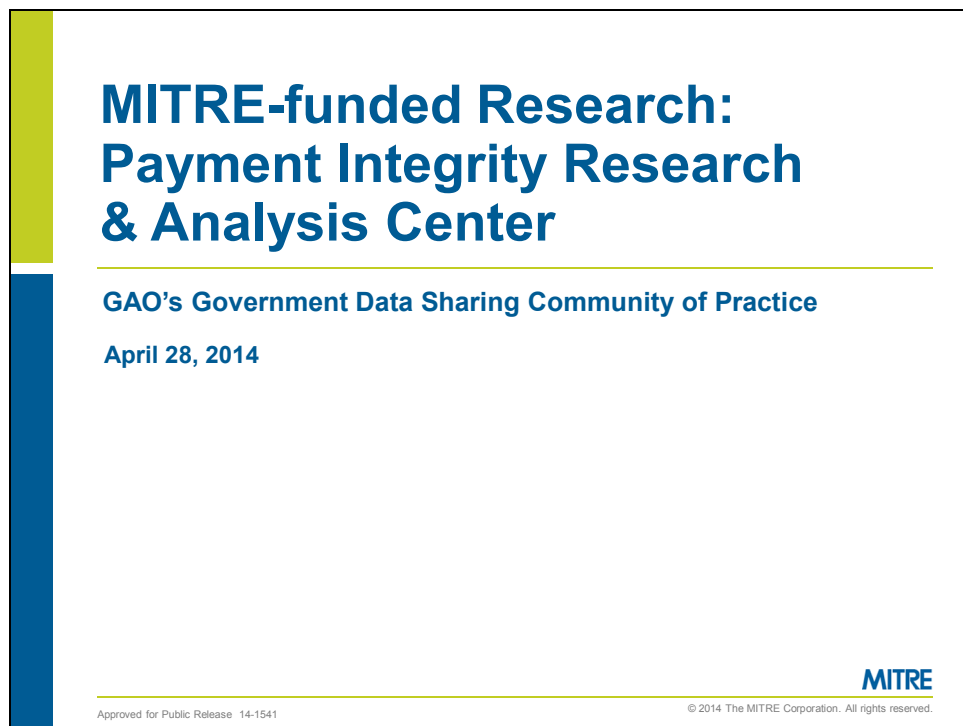## Gordon Milbourn – Principal, The MITRE Corporation

Mr. Milbourn began his presentation by providing an overview of MITRE's role as a Federally Funded Research and Development Center, as well as MITRE's efforts in helping agencies and the oversight community combat fraud and other improper payments more efficiently and effectively through data sharing.

Mr. Milbourn also introduced MITRE's hypothesis for the Payment Integrity Research & Analysis Center that could help find ways to fight fraud more effectively and efficiently both within individual agencies and for cross-government issue areas. To do this, the center could collect

data from relevant agencies, as well as commercial data sets and tools, to create a one-stop-shop for users to leverage a rich environment of data and capabilities. It would be particularly beneficial, Mr. Milbourn, said for use on cross-government efforts such as disaster recovery, where it could provide valuable information and analytics to agencies responding to disasters, such as the Department of Housing and Urban Development (HUD), the Federal Emergency Management Agency (FEMA), and the Small Business Administration (SBA). Additionally, the center could house only metadata so that it would not need access to personally identifiable information, which could be stored offline or protected by the agency. This would also ensure the anonymity of data.

In addition to providing data sharing and analysis capabilities, Mr. Milbourn said the center could catalog and assess current analytic approaches; recommend improvements; and provide advanced analytic capabilities, as well as training to help users understand the best approach to solve their specific improper payments problems.

MITRE-funded Research: Payment Integrity Research & Analysis Center

GAO's Government Data Sharing Community of Practice

April 28, 2014

MITRE

Approved for Public Release 14-1541   © 2014 The MITRE Corporation. All rights reserved.

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Bottom Line Up Front

- **As an FFRDC, MITRE conducts extensive research**

- **Overarching fraud research theme:** *Find ways to help agency officials, the accountability community, and law enforcement combat fraud and other improper payments more effectively within the constraints of existing legislative, budgetary, technical and cultural impediments to data sharing*

- **Identifying fraud and other improper payments (esp. cross-government) faces a number of challenges**

- **Can many of these challenges be addressed by means of a government-wide Payment Integrity Research & Analysis Center?**

Approved for Public Release 14-1541 **MITRE**

Numerous Issues Present in Domain

- **Fraud attacks – numerous, diverse, evolving**
- **Defensive challenges**
  - Unaligned fraud prevention / detection policies and practices
  - Incomplete, vague, inaccurate data
  - Constraints on sharing data
  - Lack of network interoperability
  - Inadequately safeguarded information
- **Defensive needs**
  - Advanced collaboration and analytic capabilities
  - Efficiency

Stopping fraudulent payments before they get out the door is critical, but has to be balanced with getting proper payments out the door, on time

Approved for Public Release 14-1541 **MITRE**

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.
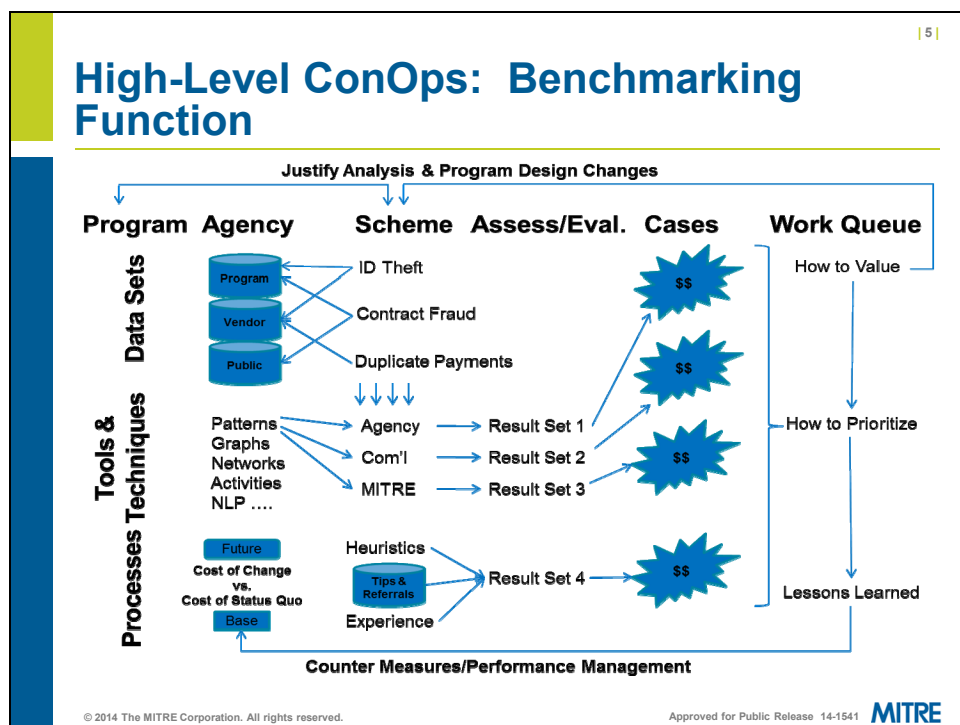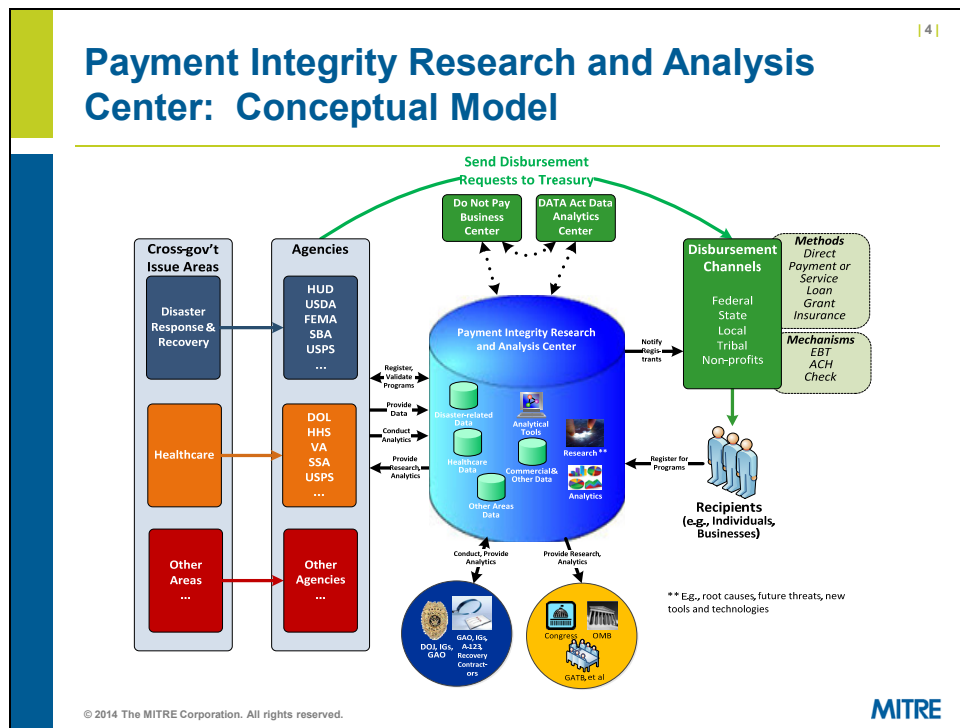
Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Selected Questions from the Moderator and Audience

Question - As agencies sort through various vendors, do you have advice on how to navigate the landscape of analytic tools?

Mr. Milbourn stated that MITRE is in the process of identifying and cataloguing more than 70 vendors to determine their capabilities. This will help develop a rich understanding of available tools and assist users with selecting the best option for their needs.

Mr. Jones said the key to developing a successful analytics program is to determine the requirements and envision the desired end result. Instead of falling into the trap of selecting the most sophisticated tool, Mr. Jones recommended asking the right questions to ensure the appropriate product is chosen.

Question - In addition to data quality, false positives present challenges for data analytics. What are some ideas to mitigate the amount of false positives so time is not wasted chasing the wrong person?

Mr. Jones said that data analytics can help investigators find false positives faster. For example, what would normally take 2 weeks to investigate and conduct surveillance on might take 2 hours using data analytics. Additionally, successful cases should be measured by whether the lead accurately identified fraud, and not whether it resulted in an investigation.

Mr. Schweikert echoed Mr. Jones' observation by stating that feedback is the most important aspect to measuring results, not the number of prosecutions. False positives are inevitable to a certain extent, but data analytic tools can help narrow the results and improve the level of certainty. It is also critical to conduct a cause analysis to determine if the right definition was used.

Question - How do you balance rapid payments with proper payments, particularly when there is pressure to get payments out quickly?

Mr. Schweikert responded that rapid payments are a high priority for the IRS and they tend to get a lot of pressure to avoid withholding payments. However, they have recently started receiving more support to slow down payments to ensure the right people are getting paid. As a result, the IRS is trying to develop an intermediate step between processing quick turnarounds and holding payments that would allow time to identify false returns while maintaining timeliness of proper payments.

Mr. Jones agreed with Mr. Schweikert that transparency is key to culture change and recommended presenting results in a nonthreatening manner and finding early adopters to help momentum. It is also important to translate this issue into terms taxpayers and businesses can understand, so that they realize the benefit of waiting an extra day or two if it ultimately means saving taxpayer money.

Mr. Milbourn stated that another challenge to balancing rapid payments with proper payments is a lack of visibility across agencies. For example, with disaster-relief programs, one agency does

not have the means to look across other disaster-relief agencies to determine whether an applicant has, either intentionally or unknowingly, applied for benefits under multiple programs. MITRE hopes to mitigate this by taking steps to catalog successful means-based testing across agencies.

Panel 3: Data Analytic Techniques to Identify and Prevent Fraud

Moderator
Joah Iannotta, Assistant Director, Forensic Audits and Investigative Service, U.S. Government Accountability Office

Panelists
- Stephen Conway, Audit Director, Office of the Inspector General, U.S. Department of Health and Human Services
- Alexandra Habershon, Program Coordinator, Integrity Vice Presidency, The World Bank
- David Partridge, Special Agent, Office of the Inspector General, U.S. Department of Defense
- Della Whorton, Assistant Director, Audit Liaison and Reports, Recovery Accountability and Transparency Board

Presentations

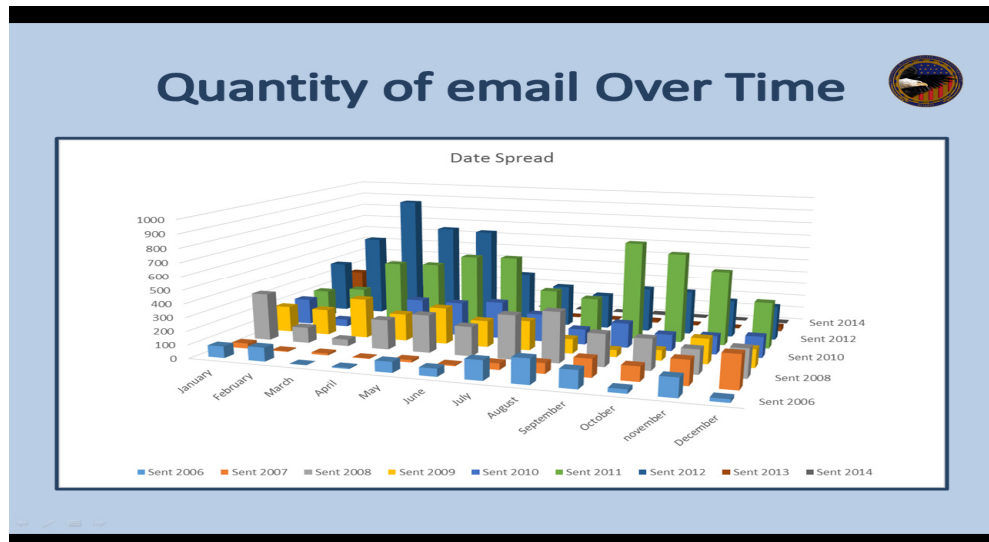Della Whorton - Assistant Director Audit Liaison and Reports, Recovery Accountability and Transparency Board

Ms. Whorton discussed two techniques used by the Recovery Accountability and Transparency Board (the Board) to analyze unstructured text data.
- Ms. Whorton first described an effort to analyze e-mail. One step in this process is to identify the last string, so that Board staff did not analyze the same e-mail chain over and over again. Board staff also developed graphics to help them understand relationships between individuals. For example, Ms. Whorton said a link diagram was developed that shows emails between the target and other subjects. The thickness of the links in the diagram indicates the frequency of contact between the subject and target.
- In her second example, Ms. Whorton described analysis related to single audits that was done in partnership with the Federal Audit Clearinghouse, which operates on behalf of the Office of Management and Budget. Board staff created cross walks to link data contained in PDF documents to other various unstructured data sets to create a structure. Key words were automatically linked in source PDFs.
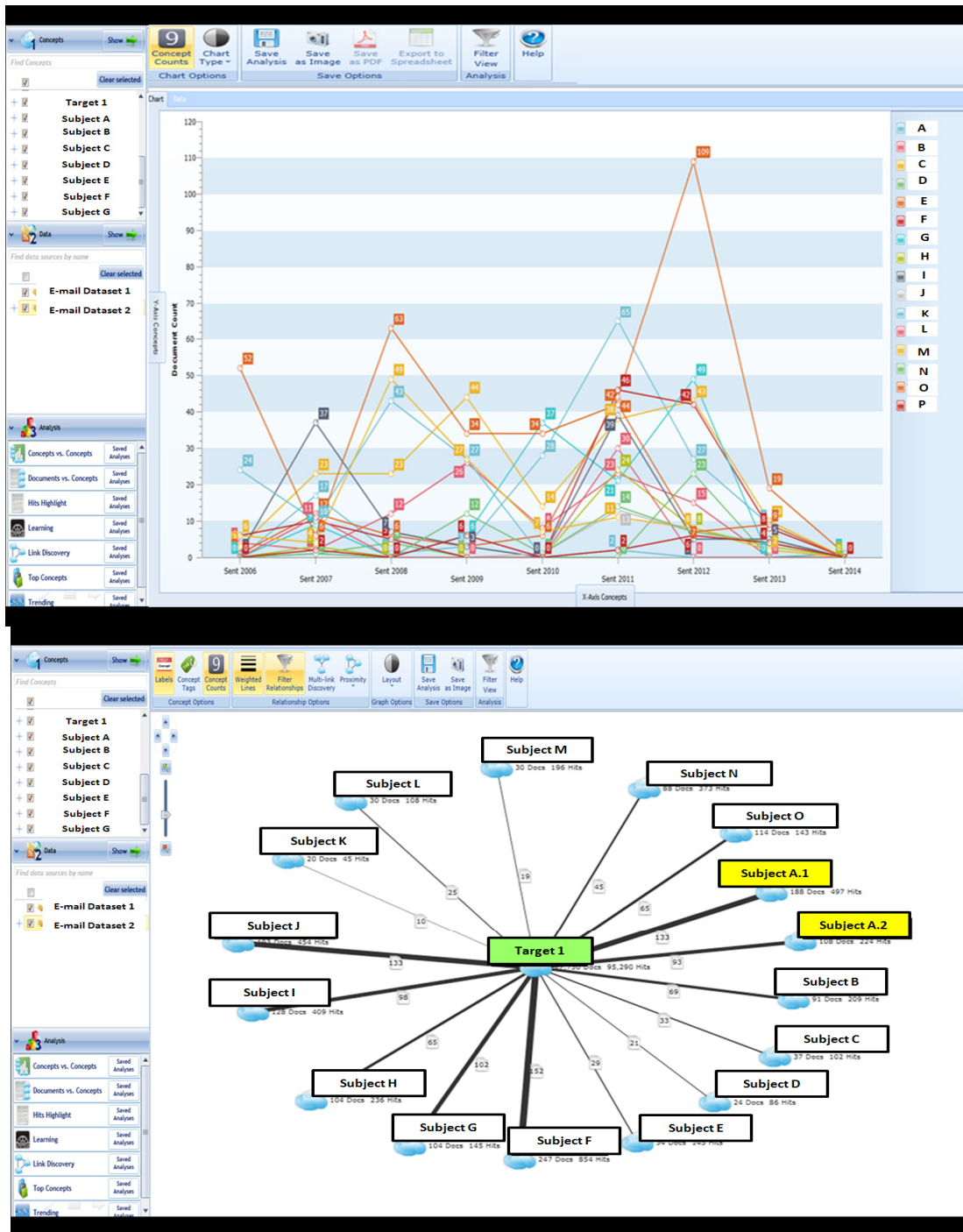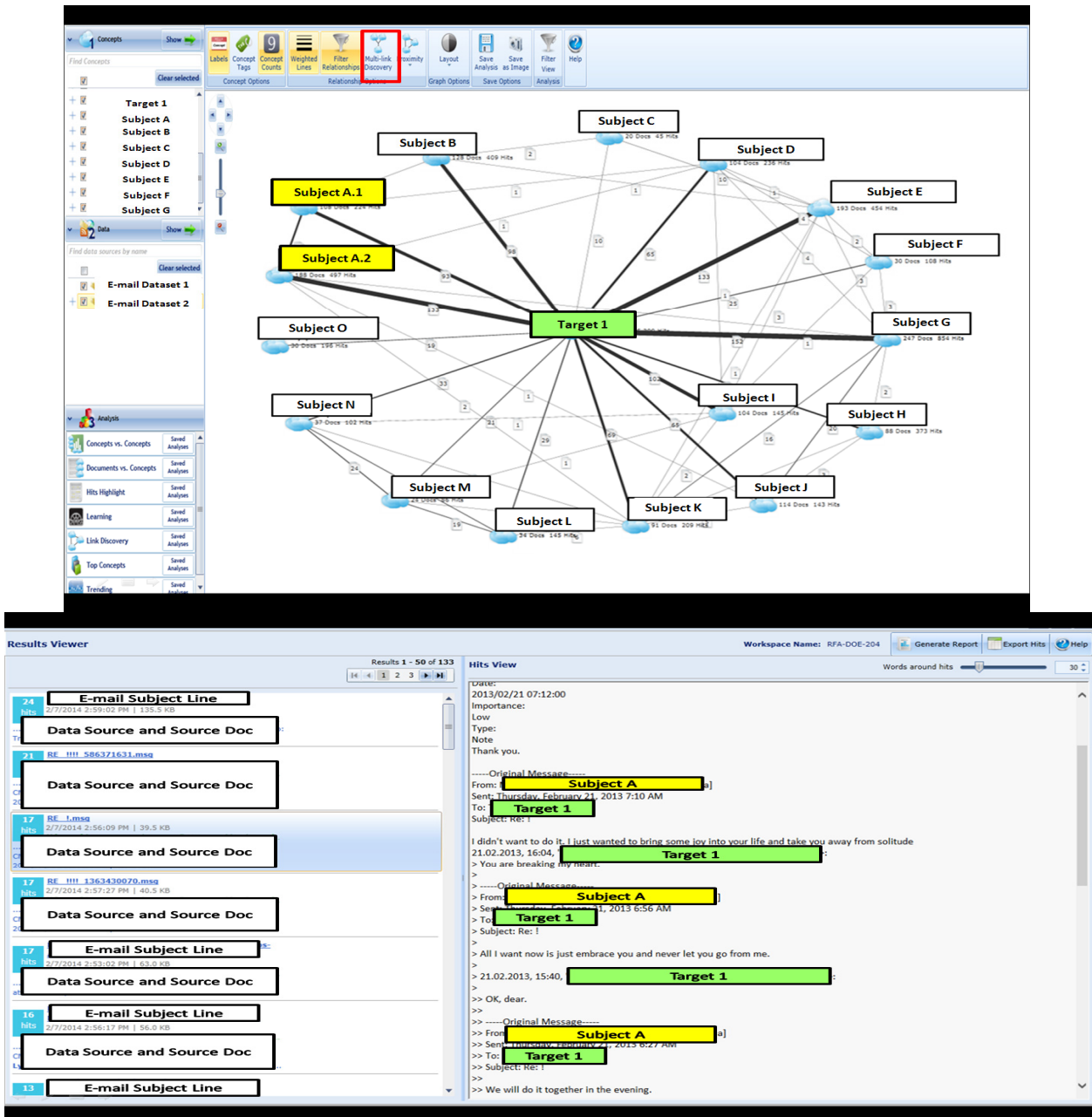
Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.
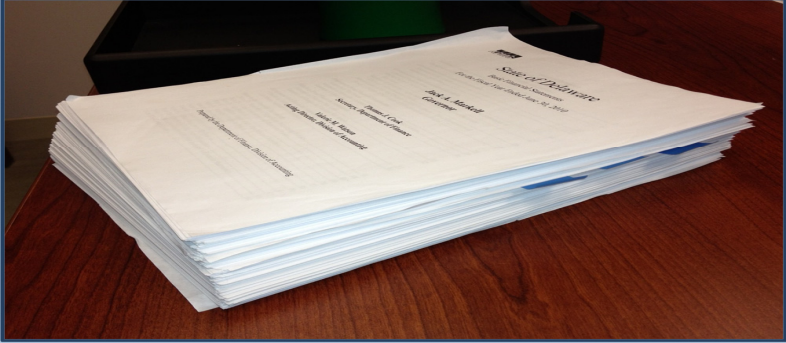
Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.
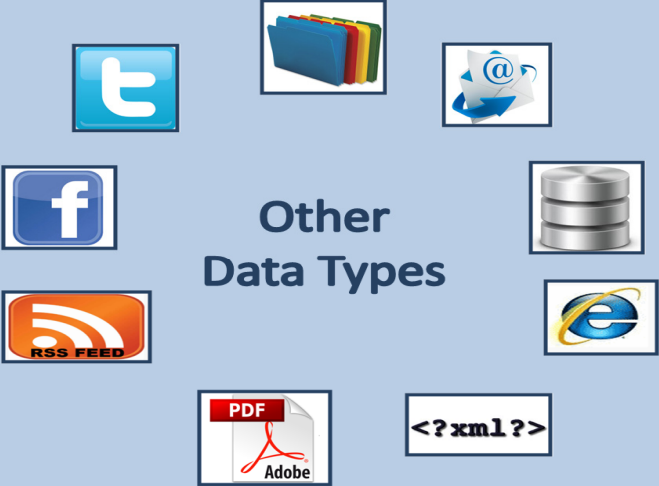
Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

= Crosswalk Analysis

Document Path to PDF Reports	Aiding and Abetting	Indicted	Lacked Adequate Documentation	Program Fraud	Document Path to Summary Table	A133 Summary Table AMOUNT EXPENDED	A133 Summary Table AUDITYEAR	A133 Summary Table CFDA	A133 Summary Table CITY	A133 Summary Table STATE	A133 Summary Table FINDINGREFNUMS
\\fagovvfs\census	1	1	1	1	db://A133SQLDSN.`	$ 156,192	2009	16.589	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\census	1	1	1	1	db://A133SQLDSN.`	$ 57,934	2009	16.587	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\census	1	1	1	1	db://A133SQLDSN.`	$ 223,347	2009	16.557	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\census	1	1	1	1	db://A133SQLDSN.`	$ 105,523	2009	16.59	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\census	1	1	1	1	db://A133SQLDSN.`	$ 86,412	2009	93.671	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\census	1	1	1	1	db://A133SQLDSN.`	$ 8,606	2009	16.806	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\census	1	1	1	1	db://A133SQLDSN.`	$ 73,792	2009	16.588	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\census	1	1	1	1	db://A133SQLDSN.`	$ 2,759	2009	16.587	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\census	1	1	1	1	db://A133SQLDSN.`	$ 39,984	2009	16.557	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\census	1	1	1	1	db://A133SQLDSN.`	$ 6,723	2009	16.587	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\census	1	1	1	1	db://A133SQLDSN.`	$ 3,966	2009	16.524	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\census	1	1	1	1	db://A133SQLDSN.`	$	2009		MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\fresh	1	1	1	1	db://A133SQLDSN.`	$ 156,192	2009	16.589	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\fresh	1	1	1	1	db://A133SQLDSN.`	$ 57,934	2009	16.587	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\fresh	1	1	1	1	db://A133SQLDSN.`	$ 223,347	2009	16.557	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\fresh	1	1	1	1	db://A133SQLDSN.`	$ 105,523	2009	16.59	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\fresh	1	1	1	1	db://A133SQLDSN.`	$ 86,412	2009	93.671	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\fresh	1	1	1	1	db://A133SQLDSN.`	$ 8,606	2009	16.806	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\fresh	1	1	1	1	db://A133SQLDSN.`	$ 73,792	2009	16.588	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\fresh	1	1	1	1	db://A133SQLDSN.`	$ 2,759	2009	16.587	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\fresh	1	1	1	1	db://A133SQLDSN.`	$ 39,984	2009	16.557	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\fresh	1	1	1	1	db://A133SQLDSN.`	$ 6,723	2009	16.587	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\fresh	1	1	1	1	db://A133SQLDSN.`	$ 3,966	2009	16.524	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\fresh	1	1	1	1	db://A133SQLDSN.`	$ 2,324	2009	16.017	MISSION	SD	09-1 09-2 07-1 02-1 02-2
\\fagovvfs\census	0	0	0	0	db://A133SQLDSN.`	$ 103,086	2008	16.54	BENTON	IL	2008-1 TO 2008-4
\\fagovvfs\census	0	0	0	0	db://A133SQLDSN.`	$ 238,351	2008	14.228	BENTON	IL	2008-1 TO 2008-4
\\fagovvfs\census	0	0	0	0	db://A133SQLDSN.`	$ 15,565	2008	97.042	BENTON	IL	2008-1 TO 2008-4
\\fagovvfs\census	0	0	0	0	db://A133SQLDSN.`	$ 100,190	2008	16.71	BENTON	IL	2008-1 TO 2008-4

Results Viewer   Workspace Name: Census   Generate Report   Export Hits   Help

Results 1 - 4 of 4

Hits View   Words around hits   57

01546743.pdf
21 hits — Data Source

01546760.pdf
21 hits 9/21/2011 2:01:03 PM | 1.5 MB — Data Source

01467385.pdf
21 hits — Data Source   .pdf

01467385.pdf
21 hits 7/12/2011 7:10:19 AM | 1.5 MB — Data Source

Less: Accumulated Depreciation
Total
5. COMMITMENTS AND CONTENGENCIES
Litigation
$56,323.90
257,434.65
(152,130.91)
$161,627.64
The former Executive Director of the Organization was indicted by a federal grand jury on July 6, 2011, for Program Fraud and Aiding and Abetting. She appeared before the courts on July 19, 2011 and pled not guilty to the indictment. The charge is merely an accusation, and the former Executive Director is presumed innocent until and unless proven guilty. The financial statements do not include accrual or provisions for loss contingencies that may result from these proceedings.
While the outcome of the above noted proceedings cannot be predicted, the Organization feels that any settlement or judgment would
..."

Document Excerpt: [Page 39 Paragraph 30]
"... 2011

Auditee
STATUS OF PRIOR YEAR INTERNAL CONTROL
FINDINGS, COMPLIANCE FINDINGS,
AND QUESTIONED COSTS
DECEMBER 31, 2010
FINDING 09-2 - LACK OF SUPPORTING DOCUMENTATION
CRITERIA:
A-87 and all contracts administered by the Organization require that costs be documented to include appropriate authorization, invoices and time records to support payroll costs.
CONDITION:
Many expenditures lacked adequate documentation, including purchase orders, invoices, and account code information.
QUESTIONED COSTS:
$60,586.47
EFFECT:
Possible misuse of federal, state, and local funds
RECOMMENDATION:
Controls need to be in place to ensure checks are not released until all proper support is attached and in place.
MANAGEMENT RESPONSE:
The current management   **Auditee**   the documentation of financial transactions. Because of this the Financial Policies and Procedures were reviewed at the end of 2010 and in
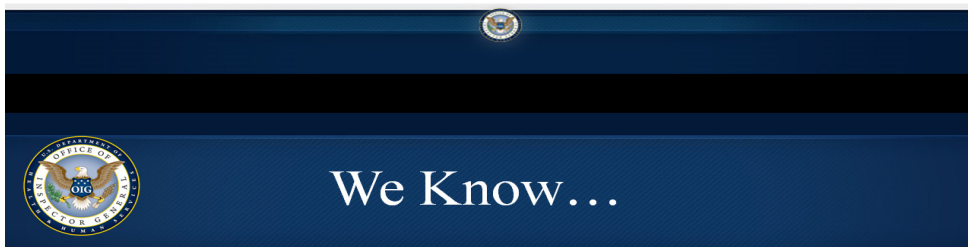..."

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Document Path to PDF Reports	Aiding and Abetting	Indicted	Lacked Adequate Documentation	Program Fraud	Document Path to Summary Table	A133 Summary Table AMOUNT EXPENDED	A133 Summary Table AUDITYEAR	A133 Summary Table CFDA	A133 Summary Table Program Name	A133 Summary Table FINDINGREFNUMS
Group 4\01467385.pdf	1	1	1	1	db://A133SQLDSI	$156,192.00	2009	16.589	RURAL DOMESTIC VIOLENCE	09-1 09-2 07-1 02-1 02-2
Group 4\01467385.pdf	1	1	1	1	db://A133SQLDSI	$57,934.00	2009	16.587	XXXX COALITION	09-1 09-2 07-1 02-1 02-2
Group 4\01467385.pdf	1	1	1	1	db://A133SQLDSI	$223,347.00	2009	16.557	ABC COALITION	09-1 09-2 07-1 02-1 02-2
Group 4\01467385.pdf	1	1	1	1	db://A133SQLDSI	$105,523.00	2009	16.59	ENCOURAGE ARREST POLICIES AND ENFORCEMENT OF PROTECTION ORDERS PROGRAM	09-1 09-2 07-1 02-1 02-2
Group 4\01467385.pdf	1	1							VIOLENCE PREVENTION	09-1 09-2 07-1 02-1 02-2
Group 4\01467385.pdf	1	1	1	1	db://A133SQLDSI	$8,606.00	2009	16.806	TRANSITIONAL HOUSING	09-1 09-2 07-1 02-1 02-2
Group 4\01467385.pdf	1	1	1	1	db://A133SQLDSI	$73,792.00	2009	16.588	VIOLENCE AGAINST WOMEN GRANTS	09-1 09-2 07-1 02-1 02-2
Group 4\01467385.pdf	1	1	1	1	db://A133SQLDSI	$2,759.00	2009	16.587	OUW RECOVERY ACT	09-1 09-2 07-1 02-1 02-2
Group 4\01467385.pdf	1	1	1	1	db://A133SQLDSI	$39,984.00	2009	16.557	TECHINICAL ASSISTANCE PROGRAM	09-1 09-2 07-1 02-1 02-2
Group 4\01467385.pdf	1	1	1	1	db://A133SQLDSI	$6,723.00	2009	16.587	SAFETY FOR WOMEN FROM SEXUAL ASSAULT	09-1 09-2 07-1 02-1 02-2
Group 4\01467385.pdf	1	1	1	1	db://A133SQLDSI	$3,966.00	2009	16.524	CIVIL LEGAL ASSISTANCE	09-1 09-2 07-1 02-1 02-2
Group 4\01467385.pdf	1	1	1	1	db://A133SQLDSI	$2,324.00	2009	16.017	SEXUAL ASSAULT	09-1 09-2 07-1 02-1 02-2

= Crosswalk Analysis

Value

Unstructured text analysis saves time and resources

17

Stephen Conway - Audit Director, Office of the Inspector General, U.S. Department of Health and Human Services

Mr. Conway discussed Health and Human Services OIG efforts to conduct "single audits"— audits that focus on multiple issues. As part of these efforts, in 2005, the office began building a Medicare claims database populated using Medicare National Claims History. To emphasize the size of the data base, Mr. Conway said each Medicare claim can generate up to 750 rows of data and hundreds of millions of claims are submitted each year.

To examine the data, he developed a Structured Query Language (SQL) script that ranks hospitals according to size and location and then ranks them based on reimbursement or

overpayment history, among other things. The results of the query are forwarded to auditors for follow up. Audits have been performed on 110 hospitals. Mr. Conway said they would like to expand this approach to Medicaid and have implemented three home-health audits. They have also added more areas of risk to the SQL code.

Data Driven Medicare Audits

Stephen J Conway
DIRECTOR, DATA ANALYTICS
US DEPARTMENT OF HEALTH & HUMAN SERVICES
OFFICE OF INSPECTOR GENERAL
OFFICE OF AUDIT SERVICES

We Know…

- Medicare Providers
  - have consistently submitted improper claims that while substantial in total (both volume of claims/dollars or over an extended period of time) have never met a threshold subjecting them to an audit or investigation by law enforcement
  - have previously been found to have committed civil/criminal fraud in one area yet have improper payments in other areas

Question Posed

How can we use data to drive
how we audit,
what we audit,
and who we audit?

"Single Audit" Approach

- Evaluate every claim against Medicare laws and regulations
- Focus on high risk issues previously identified by OIG, CMS contractors, or DOJ
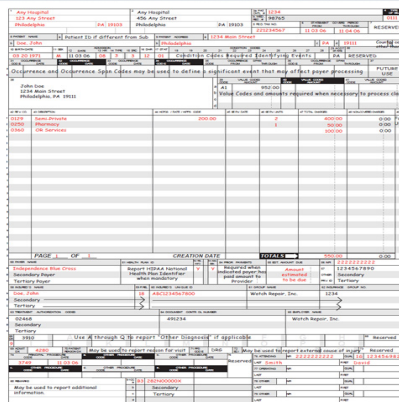- Create a baseline
- Effective
- Efficient
- Sentinel effect

OIG Data Warehouse

□ Houses claims data for
  □ Medicare Parts A, B, C, and D
□ Reference Data
  □ Enrollment (Beneficiary and Provider)
  □ Death Master File
  □ Compromised Number Checklist
  □ DEA Registrants
  □ Excluded Providers
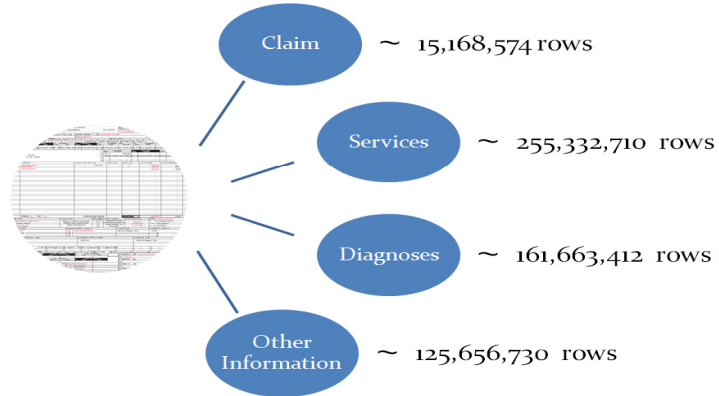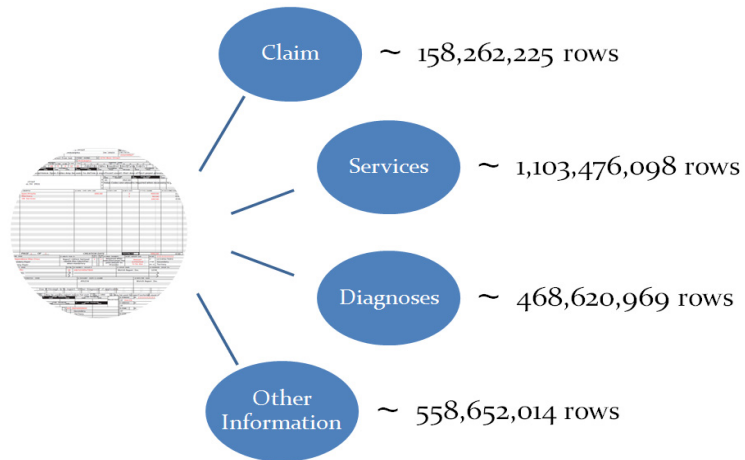 □ Medicare Overpayments

Data Driven

6

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Inpatient

Claim ~ 15,168,574 rows
Services ~ 255,332,710 rows
Diagnoses ~ 161,663,412 rows
Other Information ~ 125,656,730 rows

Outpatient

Claim ~ 158,262,225 rows
Services ~ 1,103,476,098 rows
Diagnoses ~ 468,620,969 rows
Other Information ~ 558,652,014 rows

SQL Script

- Single program involving nearly 1,000 individual SQL statements
  - Reimbursement history
  - National ranking
  - Overpayment history
  - Provider claims
  - All other claims for beneficiaries
  - Identify and quantify claims for each compliance issue
  - Download claims information for audit staff

Results

- Hospitals
  - 110 hospitals audited
  - To date we have recommended providers return $86.2 million to Medicare Trust Fund
  - $30 to $1 ROI when medical review is used

Moving Forward

- Expansion
  - Application to Medicaid
    - Data
    - Various State programs
  - Application to other provider types
    - 3 pilot Home Health audits underway
  - Additional compliance issues

Alexandra Habershon - Program Coordinator, Integrity Vice Presidency, The World Bank

According to Ms. Habershon, The World Bank is interested in data analytics because approximately $20 billion to $40 billion of development funds are lost to corruption per year. The World Bank's efforts are presently in a conceptual phase and have focused on identifying possible open data sources. Ms. Habershon said that efforts to identify data have been informed by a larger question about how data mining can be used to investigate corruption. To this end, they have explored information available in-house, such as donor contracts.

They have also looked at data sources outside The World Bank. Ms. Habershon said they have tried to find indicators of fraud by examining procurement records in national procurement websites. As part of these efforts, they have tried to pick websites that have machine-readable data and information on contract winners and losing bidders. Ms. Habershon said they have also started thinking about what suspicious bidding patterns look like and have looked at collusion between contractors in the bidding process, as well as collusion between the procuring agency and contractor and vanishing bidders.

Ms. Habershon said they also tried to foster outside-of-the-box approaches to how the data could be used by hosting a data dive, where volunteers were encouraged to "hack" the data and explore new ways for using it.

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Why invest in new tools to fight corruption?

Every year the World Bank provides more than US$ 30 billion in funding to over 140 member countries to support development projects

Every year corruption drains an estimated US$ 20-40 billion from developing countries (World Bank estimates)

Crime, corruption and tax evasion combined drained over $900 billion from the developing world in 2011 (Global Financial Integrity estimates)

"Corruption is public enemy number one"
                                  World Bank President, Jim Kim

Integrity Vice Presidency
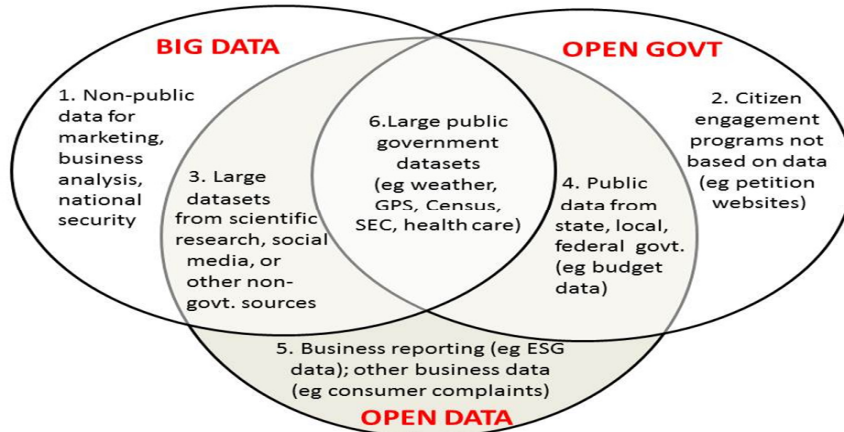The World Bank

THE WORLD BANK
IBRD • IDA

Can we use data analytics to detect corruption, collusion or fraud?

A. Are there indicators of fraud or corruption in project financial management records?

B. Can we detect patterns of collusion or corruption in procurement records?

C. Are we doing business with companies related to debarred entities? With shell companies?

Integrity Vice Presidency
The World Bank

THE WORLD BANK
IBRD • IDA

Source: Joel Gurin, Big Data Now.

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.
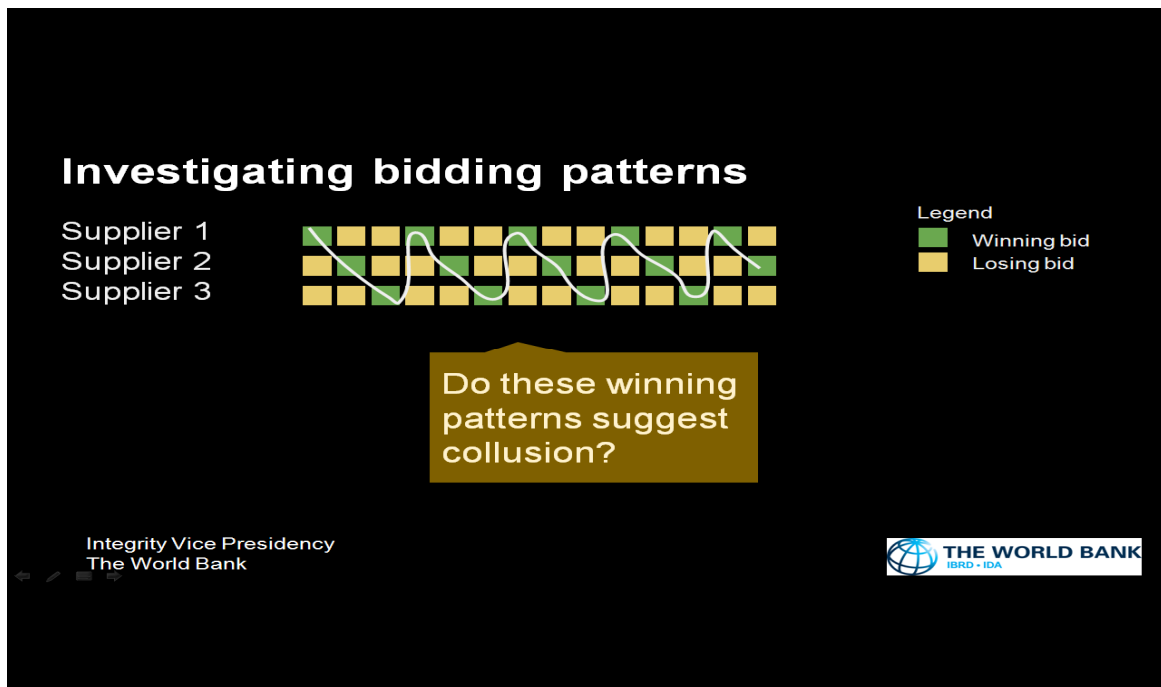

30

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.
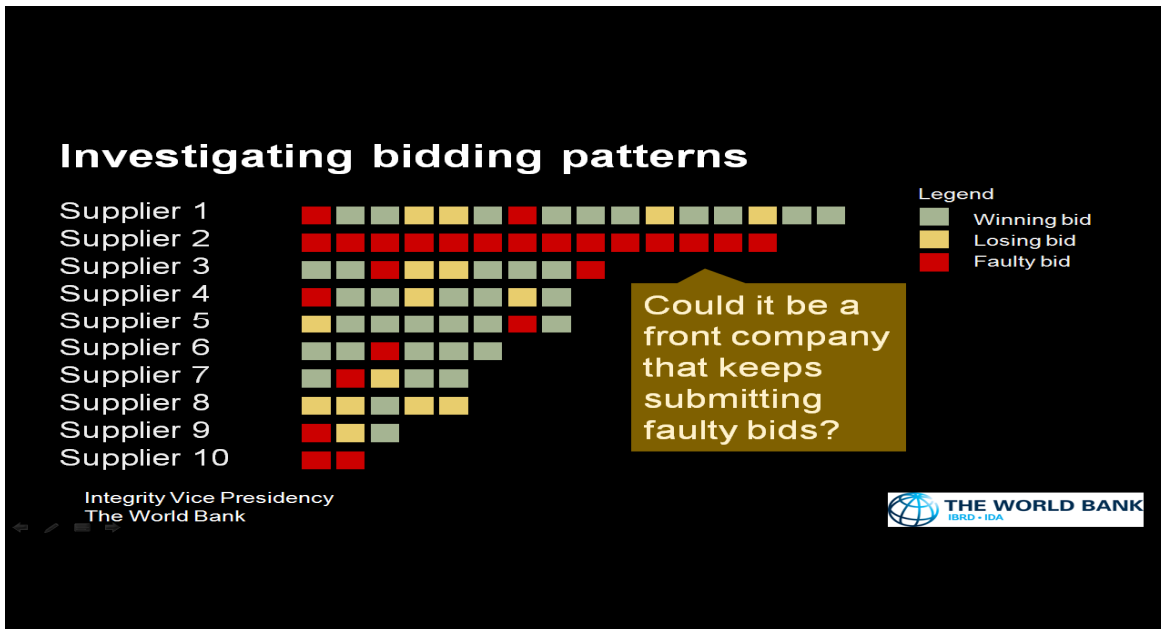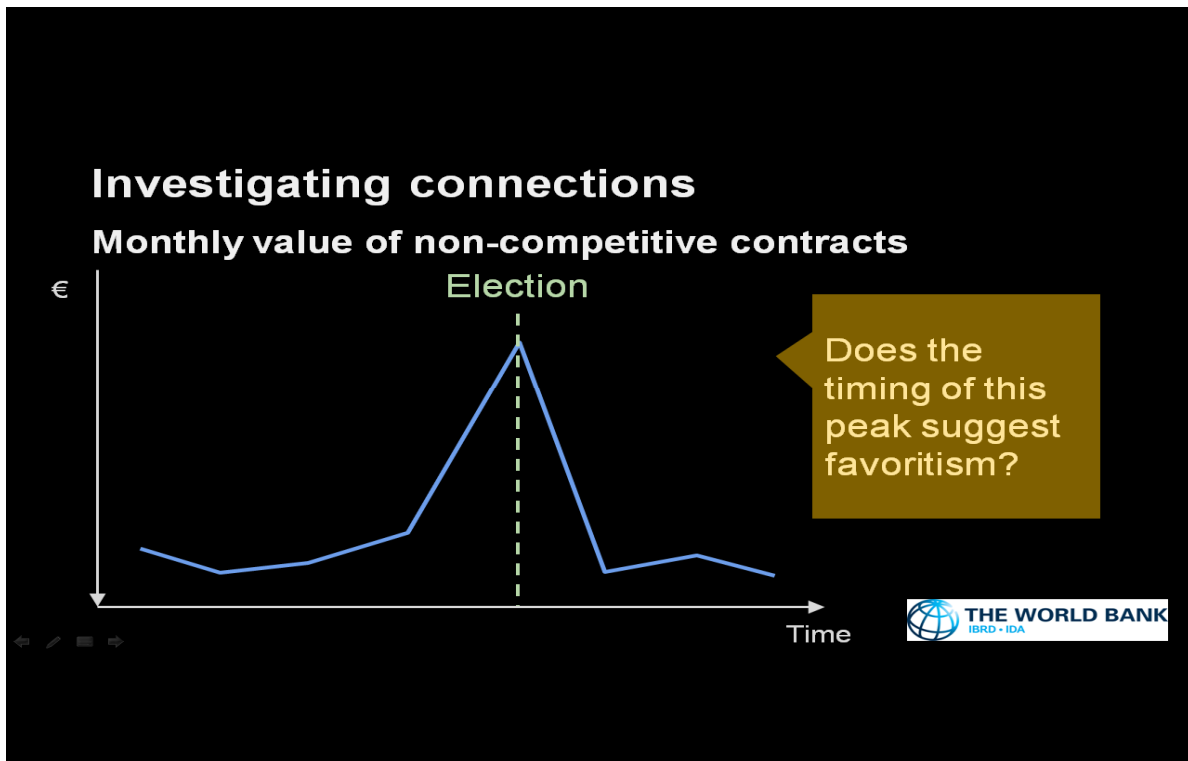

Alexandra Habershon
ahabershon@worldbank.org

Elisabeth Wiramidjaja
ewiramidjaja@worldbank.org

Steve Burgess
sburgess@worldbank.org

Integrity Vice Presidency
The World Bank

Slides by Francis Gagnon

THE WORLD BANK
IBRD • IDA

<u>David Partridge - Special Agent, Office of the Inspector General, U.S. Department of Defense</u>

Mr. Partridge described his previous experience as an investigations manager at a financial institution, where he led efforts to develop a system that was intended to predict and provide early detection of fraud committed by current and former employees. To inform these efforts, his team tried to think like "bad guys" about common types of crimes, such as self-dealing, account take-over, and selling credit information. Mr. Partridge said they also interviewed fraudsters to learn about internal controls weaknesses.

His team developed a list of red flags/indicators of crimes and conducted an internal assessment to determine whether their organization's information technology and telephony had the capacity to collect data associated with the red flags/indicators. The beta version of the system was used to catch "low-hanging fruit" such as Customer Service Representatives accessing customers' accounts with no note indicating why the account was accessed. In the second phase of system development, Mr. Partridge said they began looking for other fraud indicators such as Customer Service Representatives changing late fees. His team also ran several investigations with the Secret Service and U.S. Postal Inspection Service to catch employees who requested credit cards using customers' identities.

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

CURRENT TECHNIQUES IN DATA ANALYTICS:

SUCCESS STORIES AND PROMISING TOOLS

Special Agent David Partridge
DoD Office of Inspector General

April 28, 2014

Vision

Senior Corporate Investigations Manager
1995-1999 – Financial Institution

• Create a system that was **designed to predict and provide early detection of fraud** committed **internally by current and former employees**.

• When **suspected fraudulent activity detected, alerts would be sent to analysts** for follow-up with customers and validate the suspected suspicious activity.

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Process – Phase 1

DEVELOPMENT

- Analysis of **past and current internal fraud** committed by **current and former employees.**

- Examined the **most common types of crimes** most likely to be **committed internally by current and former employees,** the **red flags/indicators** associated with those types of crimes such as <u>self-dealing; account takeover; identity theft; and selling of credit card numbers from dormant accounts or high credit limit accounts.</u>

Process – Phase 1

DEVELOPMENT

- Conducted an **internal assessment** by our IT and telephony team members to determine whether we had the **internal capabilities and capacity** with respect to **in-house stored data to capture any information** that would closely **align with the red flags/indicators.**

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Process – Phase 1

INITIAL CONSIDERATIONS AND LIMFACS

➢ How to **reduce the rate of false positives** and **increase overall accuracy** regarding suspected criminal behaviors.

➢ Ability to **update database** and **refine** the predictive **model.**

➢ How to **identify new fraudulent schemes.**

➢ **Process took approximately 4-6 months** (Project team met on an average of 2 times each week.

Process – Phase 2

IMPLEMENTATION

❖ Programmed data into the system and put out **Beta version to catch the more simple types of fraud** such as Customer Service Representative (CSR) opening customer account with no valid incoming call or note reflecting the customer inquired about the account.

❖ Next, we **integrated behaviors** associated with CSR's **opening multiple accounts (high credit limit or dormant accounts**).

Process – Phase 2

IMPLEMENTATION

❖ As **process** was **refined**, we then looked at **expanding other fraud indicators** such as CSR changing APR, late fees, over limit fees which was cross referenced against **current and former employee database** that included **specific data on each employee (name, former names, current and former supervisors, references, emergency points of contact, addresses, phone numbers, etc.)**

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Process – Phase 2

IMPLEMENTATION

❖ Further refinement included the **aspect of any changes requesting lost, stolen cards and new PIN numbers within 50 mile radius**.

❖ For example, **all zip codes, phone numbers (current and prior) addresses and emergency point of contact information for all current and former employees was programed into the system.**

Process – Phase 2

IMPLEMENTATION

❖ Any time a request came in for a lost card, PIN change, access checks or something similar happened, **the program would compare the requestor's address to determine if it was similar or in close proximity to the employee** (to include former employees) and send an alert down to an analyst for follow-up.

Process – Phase 2

IMPLEMENTATION

❖ The analysts would pull the CSR's notes to **determine any reference to an incoming call and if information reflected no valid reason for accessing the customer's account,** the analyst would then **call the account holder** on file and **verify that the customer had in fact called in about their account.**

❖ If the customer did not, we would **flag and monitor the account** and if the individual called in again, the call would be forwarded to one of our analysts and they would portray that they were a CSR.

35

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Process – Phase 2

IMPLEMENTATION

❖ At that point, the Director of Corporate Security and the investigations staff would discuss whether we wanted to just **stop the fraud** or **continue gathering information and pursue criminal prosecution with state/federal law enforcement authorities.**

❖ Typically, we looked at factors such as the **employee's position and length of employment, the dollar amount of fraud committed, the length of time the internal fraud had been occurring** and make a determination on involving the local law enforcement or terminating the employee.

Process – Phase 2

IMPLEMENTATION

❖ Ran **several operations with the Secret Service, US Postal Inspection Service (USPIS),** etc. In those cases we would get with embossing and make a bogus card.

❖ The **card would be hand-carried** by one of our investigators and would **be tracked from when it left the facility to the caller's residence** and working with USPIS and local law enforcement, we could **conduct surveillance** on the individual and **arrest them once they opened their mail box and had possession of the bogus credit card.**

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Process – Phase 2

IMPLEMENTATION

❖ The **system screened calls coming in from individuals with similar names as employees asking for PIN changes, new cards or that were applying for a car loan.** Had several instances of identity theft where former employees applied for and were granted a loan based on a similar variance in name.

Process – Phase 2

IMPLEMENTATION

❖ The INDETECT system **would filter and link an incoming call from the account holder and check the CSR's activity based on that incoming call.**

❖ The **overall objective** of the program was designed to **monitor suspicious activity** associated with customer's accounts or CSR activity.

Example

❑ If a CSR opened my account from their computer and there was **no valid reason for anyone to look at my account,** an **alert would flow down** to one of the analysts.

❑ They would then **call me** and ask if **I called** into Customer Service **inquiring about my account.** The system also monitored CSR activity when several accounts were opened and the CSR seemed to be just looking at accounts (trolling), especially if they had been dormant accounts or high credit limit accounts.

❑ Bottom line, the CSR **should not be opening** and viewing accounts unless there was a **valid reason** for doing so.

Application Within the Federal Government

❖ Start with **initial data, run beta test** and as the system starts detecting fraud, then **expand into inputting more data to catch additional fraud**.

❖ Could be implemented in the detection of duplicate payments and/or vendor fraud.

Application Within the Federal Government

RELEVANT DATA REQUIRED

- Full Address(s) of Vendor (variations of address, number and street, especially PO Boxes)
- Full Name of Vendor (variations of name, first and last)
- Check Numbers written out prior to vendor
- Amounts of checks written out to vendor
- Vendor's CAGE Code
- Tax ID Number
- Vendor's Bank and Account Number

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

Application Within the Federal Government

RELEVANT DATA REQUIRED

- System should programmed to flag data and when there is a duplicate or suspected duplicate payment for follow-up.

Keys to Success

➢ Through involvement with the fraud detection team, continuously work on effective ways to **reduce the rate of false positives** and **increase overall accuracy** regarding suspected criminal behaviors.

➢ Continue to assess the need for **additional data** required to **update database** and **refine** the predictive **model.**

➢ Be creative in the identification of **new fraudulent schemes** and the **methods and tools for detection**

➢ Periodically assess the types of criminal activity that has occurred in the past/present and examine how this was allowed to occur and what red flags/indicators could be programmed into the system to detect and prevent future fraudulent activity.

Selected Questions from the Moderator and Audience

Question - Given all of the data and analytic tools available to you, how did you make decisions about what data to leverage and what analytic tool to apply?

Ms. Whorton said that finding something that is scalable is tough and requires testing. Another challenge is proving that this type of research and analytics has value.

Mr. Conway said that more is less: there are a lot of risk areas and too much data for auditors to evaluate.

Ms. Habershon said they identified the specific need first. They needed to evaluate how they identify risks and use opportunities to use open data and aggregate those sources together.

Mr. Partridge said his team started by looking inside the company: everyone else always starts by looking outside the company.

Question - Any insights into the cost and benefits of your program? What are some of the things you considered when choosing this tool or that popped up during your analysis?

Ms. Habershon said they have not yet been able to quantify the benefits of their efforts. However, Ms. Habershon said that data dives really pushed forward The World Bank's effort to find new ways of fighting corruption. During the data dives, they tried to push volunteers to develop relationships between data outside of the norm.

Mr. Conway also noted that the sentinel effect of the audit was huge and one of the best outcomes.

Ms. Whorton said that the Board is not an agency that oversees programs, but conducts analysis for OIGs. Her team's goal is to evaluate risk and develop leads that result in a positive record of investigation.

Mr. Partridge said that while there was initially some internal push back, they were able to generate results quickly.

Panel 4: Executive Perspectives on Data Sharing

Moderator
- Seto Bagdoyan, Acting Director of Audits, Forensic Audits and Investigative Service, U.S. Government Accountability Office

Panelists
- Jodi Patterson, Director, Return Integrity and Correspondence Services, Internal Revenue Service
- Dr. Brett Baker, Assistant Inspector General for Audit, Office of Inspector General, National Science Foundation
- Christina Ho, Executive Director of Data Transparency, U.S. Department of the Treasury
- Rod DeSmet, Deputy Assistant Inspector General for Audit, Office of Inspector General, U.S. Department of Agriculture

Presentations

Jodi Patterson – Director, Return Integrity and Correspondence Services, Internal Revenue Service

Ms. Patterson opened the panel presentations with a discussion of her role at the Internal Revenue Service (IRS). As the Director of Return Integrity and Correspondence Services, Ms. Patterson's work focuses on strengthening the tax system by preventing improper payments of tax refunds. In an effort to increase its prerefund fraud detection, the IRS recently decided to move its data analytics modeling in-house, in lieu of contracting out. As identity theft increases, the IRS is trying to prevent payments on identity theft returns through use of data modeling, fraud detection and prevention. In processing year 2012, the IRS prevented 3 million fraudulent refunds totaling about $20 billion.

In terms of data sharing, the IRS's biggest challenge is that legislation prevents it from sharing data unless it is related to tax administration. However, in recent years, the IRS has begun to find ways that would allow it to share more data. For example, the IRS has historically been unable to share tax information with law-enforcement agencies. However, as identity theft has grown in popularity among organized crime, the IRS has found a means to share data with law-enforcement agencies by reaching out to the victims of identity theft and obtaining consent to release all tax information for that individual to law-enforcement agencies, including any false returns.

Additionally, the IRS is able to share data across agencies when granted legislative permission. In an effort to identify and prevent phony tax returns from prisoners, the IRS sought legislation to allow the IRS and states to share prisoner data. Ms. Patterson stressed the importance of working closely with legislators when seeking legislation. Although, in previous legislation, the IRS received permission to share tax information, the legislation did not allow IRS to share certain parts of the tax return, such as the signature field, which was needed by the states to conduct handwriting analyses that could link the false return to a prisoner's handwriting. By working with legislators, the subsequent legislation included that provision.

Ms. Patterson concluded her presentation by discussing other challenges to data sharing such as agencies having the technology and other resources to use the data, and the effect of data quality on usability.
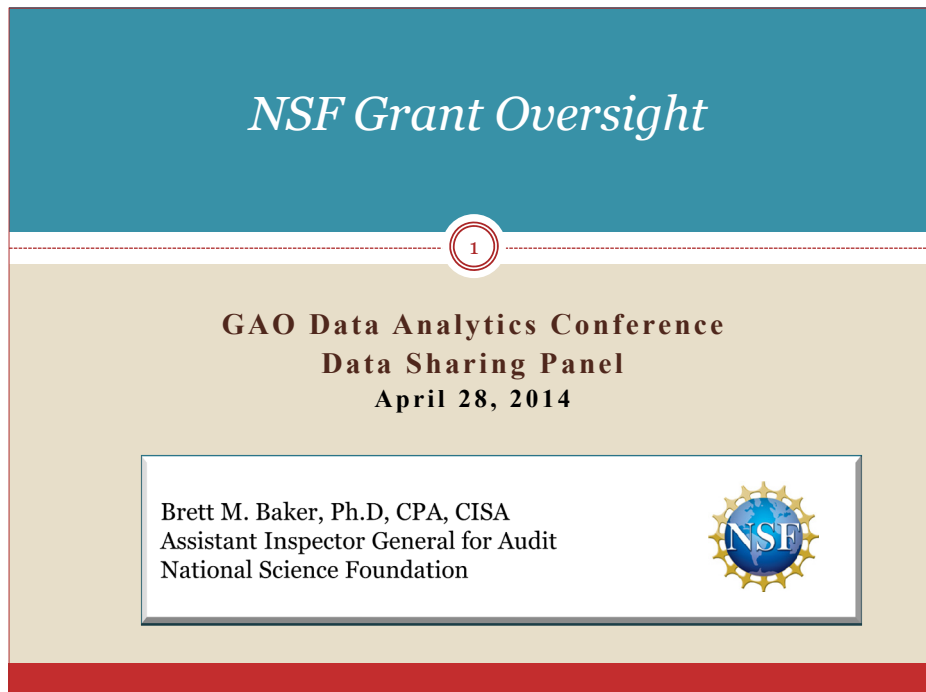
Dr. Brett Baker – Assistant Inspector General for Audit, Office of Inspector General, National Science Foundation

Dr. Baker began his presentation with a discussion of oversight challenges at the National Science Foundation (NSF). The NSF makes approximately $550 billion in awards each year. Because funds are expended at the grantee level, NSF experiences challenges in determining how the funds are actually expended. To mitigate these challenges, Dr. Baker developed a life cycle approach to oversight that entails a review of process risk from start to finish – from when the grant is awarded through the end of the project. Using data analytics, NSF then focuses on anomalous grant transactions. He has worked closely with the other OIGs and the Recovery Accountability and Transparency Board to promote this approach throughout the oversight community.

Dr. Baker further described the processes used in the life cycle approach to grant oversight. In the preaward phase, a potential award recipient is checked against the Excluded Parties List System (EPLS) to ensure a project can be funded. Then, as the award progresses, NSF reviews the expenditure burn rate to determine whether funds are expended evenly over the award period. Projects with high burn rates toward the end of the award period, which could indicate that recipients are inappropriately expending leftover money, are flagged and reviewed. At the end of the life cycle, NSF focuses on cost transfers and spend-out – for example, if money is spent after the end of the award period.

Since NSF only has visibility over the expenditure transaction itself, NSF relies on data from other sources such as agency award systems, EPLS, and Federal Audit Clearinghouse (FAC) A-133 audits, as well as recipients' financial-system records for transaction documentation needed to identify material weaknesses. Dr. Baker concluded his presentation by expressing his desire to continue expanding data sharing and collaboration with other agencies.

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

NSF   *U.S. Financial Assistance Overview*

②

- $550 billion in awards
  - 88,000 awardees and 26 Federal grant making agencies
  - Project and research, block, and formula
- Outcomes are designed to promote public good
- Challenges
  - Limited visibility of how Federal funds are spent by awardees
  - Support for funding requests much less than for contracts
- Opportunities to enhance oversight with less
  - Automated oversight

Dr. Brett Baker
AIGA. NSF-OIG

NSF   *Framework for Grant Oversight*

③

- Data analytics-driven, risk-based methodology to improve oversight
  - Identify institutions that may not use Federal funds properly
  - Techniques to surface questionable expenditures
- Life cycle approach to oversight
  - Mapping of end-to-end process to identify controls
  - 100% review of key financial and program information
  - Focus attention to award and expenditure anomalies
- Complements traditional oversight approaches
  - Techniques to review process and transactions are similar
  - Transactions of questionable activities are targeted

Dr. Brett Baker
AIGA. NSF-OIG

Note: These minutes summarize the discussion that took place at the Government Data Sharing Community of Practice meeting. The summary does not necessarily represent the views of GAO or the organizations that the discussion participants represent.

End to End Process for Grant Oversight

SOLICITATIONS · PROPOSALS · PRE-AWARD REVIEW · AWARD · CASH REQUEST · PAY/ENTITLEMENT · CASH DISBURSEMENTS · POST AWARD · AWARD CLOSE-OUT

PRE-AWARD RISKS	ACTIVE AWARD RISKS	AWARD END RISKS
•Funding Over Time •Conflict of Interest •False Statements •False Certifications •Duplicate Funding • Inflated Budgets •Candidate Suspended/Debarred	•Unallowable, Unallocable, Unreasonable Costs •Inadequate Documentation •General Ledger Differs from Draw Amount •Burn Rate •No /Late/Inadequate Reports •Sub-awards, Consultants, Contracts •Duplicate Payments •Excess Cash on Hand/Cost transfers •Unreported Program Income	•No /Late Final Reports •Cost Transfers •Spend-out • Financial Adjustments • Unmet Cost Share

D A T A   A N A L Y S I S

Dr. Brett M. Baker, 2010

Data Sources

- Internal
  - Proposals: budgets, panel scores
  - Agency award systems, recipient reporting
- External
  - Excluded Parties List System (EPLS/SAM)
  - Central Contractor Registration (CCR/SAM)
  - Federal Audit Clearinghouse (A-133 Audits)
- Recipient financial system records
  - General ledger and subsidiary ledgers
  - Effort reporting
  - Property
  - Travel and purchase card
  - Subaward monitoring
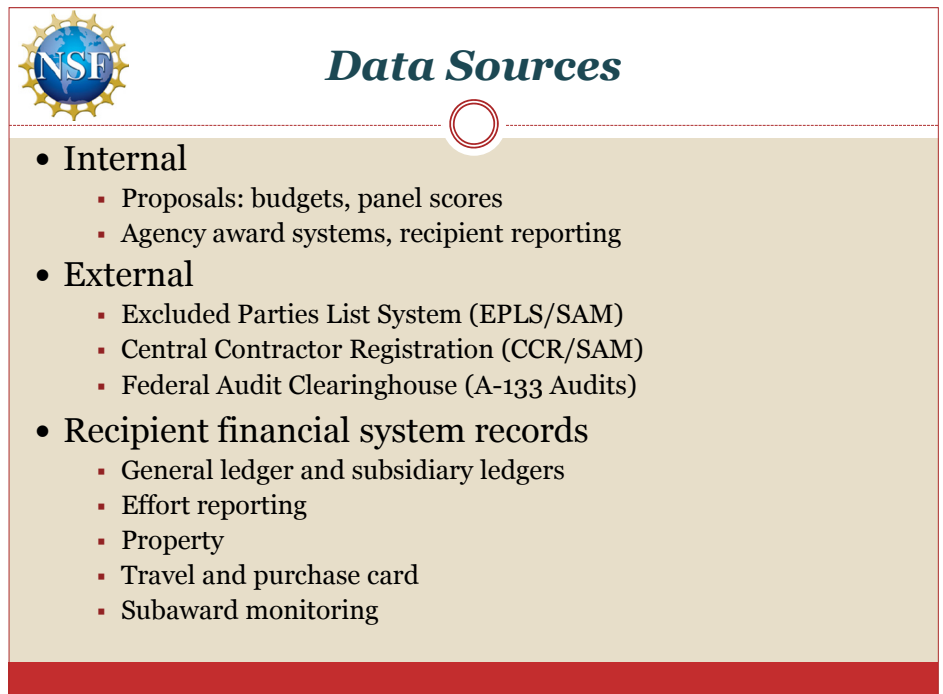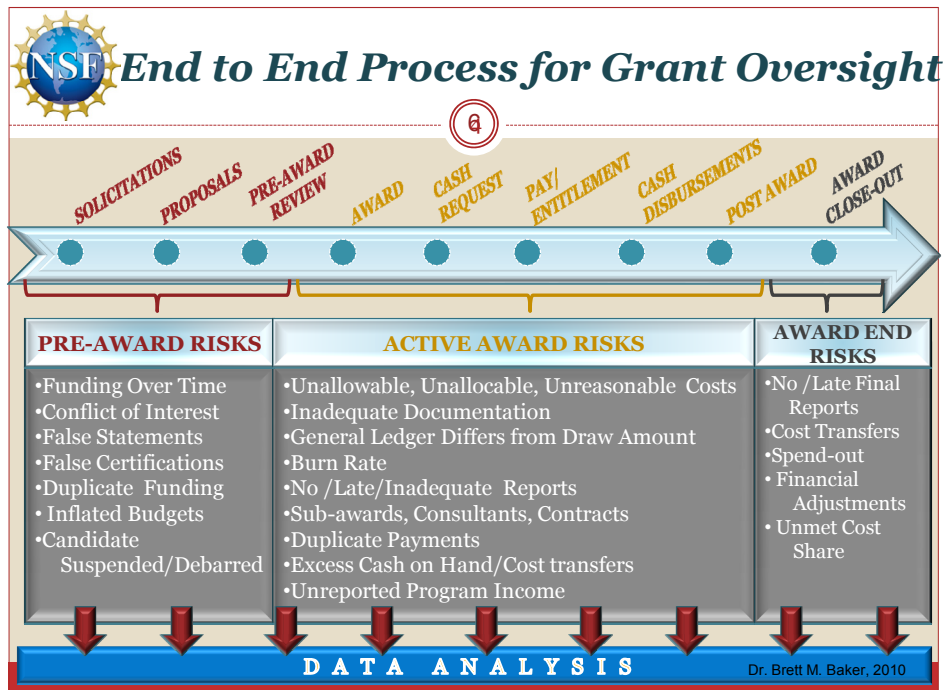
Christina Ho – Executive Director of Data Transparency, U.S. Department of the Treasury

Ms. Ho provided an overview of the Department of the Treasury's (Treasury) efforts to improve data transparency. As one of Treasury's strategic goals, the long-term vision of data transparency is to provide reliable, timely, and secure financial-management data to promote transparency, improve interoperability, and inform the public as well as other agencies. While data transparency provides the public with an understanding of how the government works, it also calls for better accessibility among agencies and usability for data analytics.

Ms. Ho cited several benefits to increased data transparency:

- Increased innovation both within the government as well as among the general public. Ms. Ho provided examples of successful innovations stemming from government data sharing efforts, such as GPS and mobile weather applications.

- Increased operational efficiencies. Specifically, increased data transparency would help the government move toward a less systems-dependent environment and improve data standards.

- A better-connected government – both internally as well as with the general public.

Ms. Ho also discussed a number of challenges associated with increasing data transparency:

- Requires a major culture change. Agencies must see the value in the vision and be willing to expose data.

- Creates challenges in determining ownership of data and responsibility for correcting known errors.

- Invites more scrutiny of government agencies.

- Invokes security and privacy concerns. Data protection would have to be balanced in a way that would keep data secure without sacrificing convenience.

Rod DeSmet – Deputy Assistant Inspector General for Audit, Office of Inspector General, U.S. Department of Agriculture

Mr. DeSmet opened his presentation with a discussion of challenges associated with data sharing, including certain challenges posed by the Computer Matching and Privacy Protection Act of 1988 (Computer Matching Act), as follows:

- Data sharing is cultural. Some agencies don't want to share and some have restrictions that will not allow them to share.

- Navigating loopholes. Someone will inevitably find a reason not to share data. For example, data involving research grants or national security concerns is often viewed by agencies as highly sensitive, which makes agencies less willing to share.

- Responsibility for data security. As part of the Computer Matching Act, accepting external data also means accepting responsibility for its security.

- Timeliness of data being shared. Some matching agreements can take a year or two to finalize before actual matching begins. Therefore, it is critical to ensure the data provided are the most recent version.

- Completeness and accuracy of data. The Computer Matching Act requires Inspectors General to review the data to determine whether they are statistically valid. Mr. DeSmet also highlighted the importance of verifying the results of matching.

- Assessing cost-benefit. Agencies need to consider how they will define the parameters concerning efforts to identify fraud and improper payments.

Mr. DeSmet also described the U.S. Department of Agriculture's (USDA) participation in Recovery.gov data work and the challenges encountered as part of its data-matching efforts. Mr. DeSmet explained that the lack of consistent standards hindered its data-matching efforts. For example, inconsistent use of hyphens and other punctuation made it difficult to match databases. Additionally, some sub-agencies reported data using a 10 percent tolerance rate, while others used 0 to 2 percent tolerance rate; thus producing inconsistent results. Although there were defined logic checks, many agencies either disregarded them or selectively used the ones they wanted.

Mr. DeSmet outlined recommendations to improve data-matching efforts:

- Develop a clear purpose. An agency should always have a clear purpose and expected outcome established before conducting data matching.

- Develop a standardized template for Computer Matching Act Agreements. Mr. DeSmet suggested that the Office of Management and Budget (OMB) could develop a standardized template to collect information and expedite the process.

- Encourage culture change. Agencies should do everything possible to ensure their programs are running effectively and efficiently.

- Develop a unique government identification number. Mr. DeSmet recommended that the government develop a uniform unique code that would help identify people and entities across government agencies.

Selected Questions from the Moderator and Audience

Question - Is there ever a risk of sharing too much information? At what point might data sharing become destructive?

Ms. Ho responded that there is a risk of the mosaic effect. The data set itself may be unclassified, but pooling individual datasets together may pose privacy or national security

concerns. However, Ms. Ho felt that it would take a while before data sharing efforts reached that point.

Ms. Patterson said that, in the public's perception, increased data sharing among government agencies may create a "big brother effect." She added that it is important to consider that agencies' missions may be at odds with each other. For example, the IRS requires every individual, whether or not they are legal residents, to file a tax return. This could create a mission conflict with the Department of Homeland Security, which is responsible for enforcing and administering immigration laws.

Mr. DeSmet said that agencies must have oversight in place to ensure that data being shared are relevant to a specific program or decision at hand. Without a clear objective, it becomes a "fishing expedition."

Mr. Baker said that, as these systems are developed, it is necessary to further consider what data are required.

Question - What is the role of technology in making data more accessible?

Mr. DeSmet responded that there is significant work to be done, technologically and culturally, before agencies are able to publish data. As systems were developed, implementation policies were left to the users, thus creating inconsistent policies and stove-piped systems. Additionally, agencies would have to undergo a cultural transformation before the data is ready to be published. Ultimately, it is important to ensure that data used and decisions made are as accurate as possible.

Ms. Ho echoed the need to make data more usable. Historically, efforts have focused on collecting as much data as possible, but the data were not necessarily useable. Today, the emphasis is on accessibility and having data that are machine readable. However, many agencies struggle with developing an effective and efficient method of extracting data out of customized systems.

Question - Given the need for data-sharing, what would be the "utopia" of data-sharing?

Ms. Ho responded that, since the stages of government spending are not intuitive and often not understood by the public, the "utopia" would be being able to provide stakeholders a complete view of the spending lifecycle from appropriation to disbursement.

Mr. Baker responded that he would like the ability to see beyond the summary-level data and down to the transaction-level data, to get a clear picture how and where the money was spent. Transaction-level data can be summarized into summary-level data, but transaction-level data cannot be extracted from summary-level data. Therefore, it is important to get transaction-level data where possible. At NSF OIG, transaction-level data shows what the funds were specifically spent on

Question - Key takeaways?

Ms. Patterson emphasized the challenges ahead with making data useable. Data should be something from which everyone can benefit and use with confidence.

Mr. Baker stated that, from a data-sharing standpoint, the opportunity to see down to the transaction level would benefit all oversight areas. Mr. Baker emphasized the need to include this type of visibility as future systems are developed.

Ms. Ho recognized that there are valid constraints and legitimate reasons to protect data. She stated that the approach moving forward should focus on being respectful and sensitive.

Mr. DeSmet stated that these types of forums provide a positive step toward increased data sharing and data use. Moving forward, Mr. DeSmet recommended that the data community also continue to keep data security and data validity in mind.