

A Spotlight on Analytics for Grant Fraud

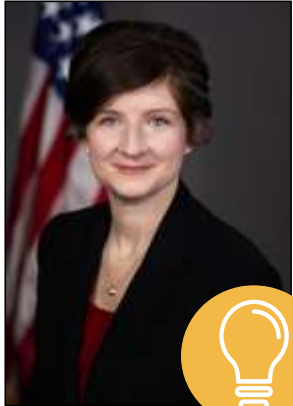


Department of Health and Human Services
Office of Inspector General
Office of the Chief Data Officer

March 4, 2020

Limited Official Use Only DHHS/OIG

OCDO: Our mission and purpose



HHS OIG OFFICE OF THE CHIEF DATA OFFICER (OCDO)

Caryl Brzymialkiewicz
Chief Data Officer

OCDO empowers OIG to use data proactively by putting data at OIG's fingertips, supporting data-driven decision-making, and helping OIG's Components deliver their oversight missions to protect the integrity of HHS programs as well as the health and welfare of the people they serve.



DATA ANALYTICS (DA)

Steve Shandy *Director of DA*

Provides OIG with the analytics needed to equip, empower, and support the agency mission of fighting fraud, waste, and abuse in HHS programs.



DATA OPERATIONS (DO)

Justin Stekervetz *Director of DO*

Leads the implementation of data infrastructure (i.e., the Integrated Data Platform) and data security to provide enterprise-wide access to self-service capabilities and data within a secure and compliant data architecture.



PLANNING, MEASUREMENT, and DELIVERY (PMD)

Michael O'Rourke *Director of PMD*

Incubates new concepts to support data-driven strategic decision-making throughout the organization and act as change agents.



OCDO: Our Results



Data Analytics completed more than **1,600 requests to support OIG's oversight efforts** since its launch, including CIGIE and NHCAA award winning projects, law enforcement takedowns, and trainings



The **Innovation and Performance Series** that spotlights best practices has been attended by over **1,000 participants in total** and all staff invited to hear key progress in the OIG Update



Data Operations expanded data governance, launched **new data capabilities** and added **new data services to the Integrated Data Platform**



Tool adoption and usage by partners has **increased** substantially—over **1 million Enterprise Dashboard views**, nearly **600 OIG analytic tool users**, and **4 versions of the Grants Analytics Portal** released

HHS OIG OCDO

Service Offerings



What is a service offering?

A service offering defines the means by which a division adds value and contributes to the vision of the OCDO.



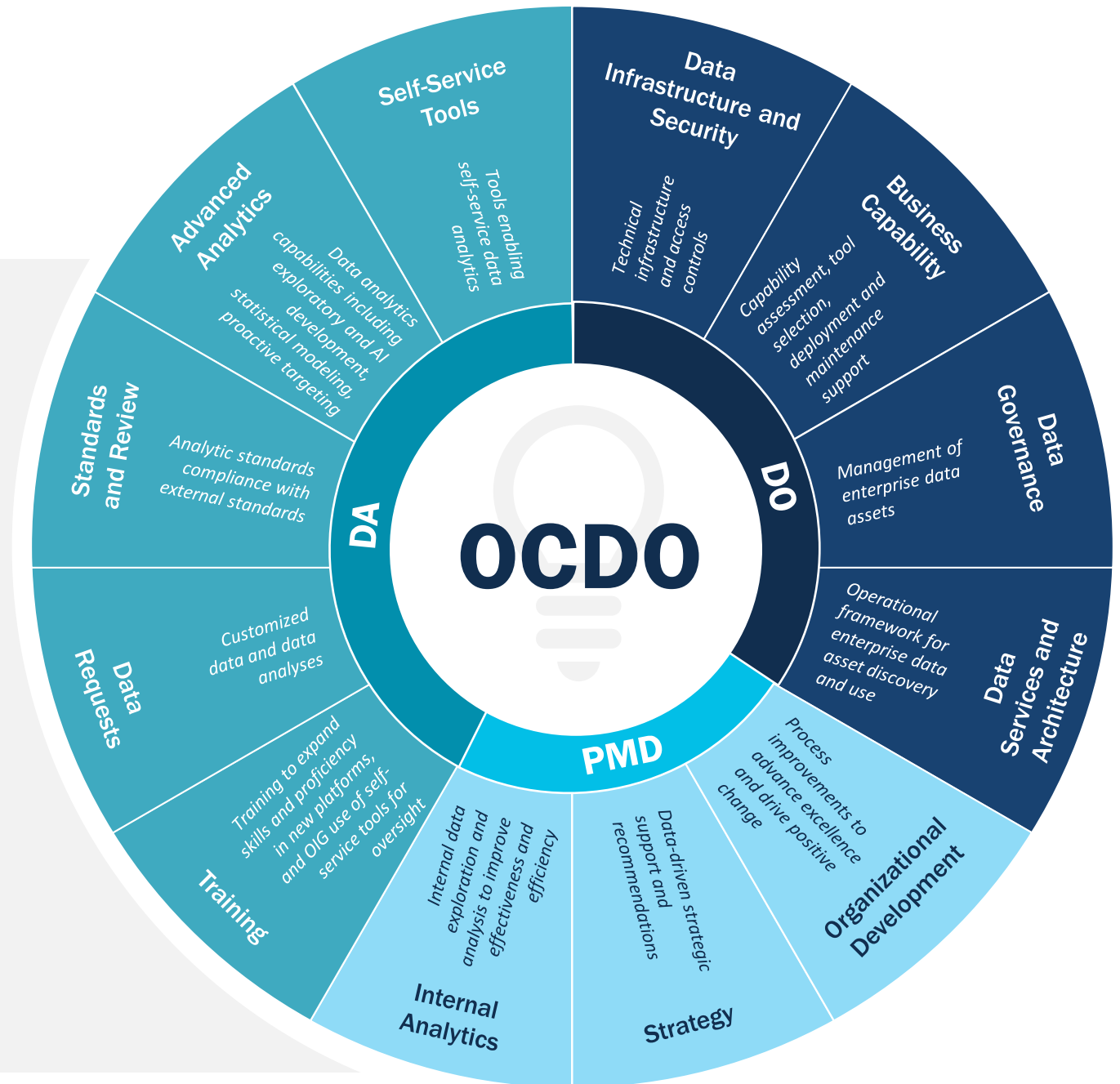
How do service offerings support HHS OIG's mission?

Providing these services are a critical step in combating fraud, waste, and abuse. These services have enabled OCDO to be a visionary for other CDOs across the government.



How does this impact our work?

Defining these services help create a common understanding of how OCDO as a whole supports OMP and OIG. This transparency also makes it easier to collaborate across divisions.



OCDO: Grants Analytics Portal (GAP)



HHS is the largest grant-making organization in the Federal Government. The [Grants Analytics Portal \(GAP\)](#) is intended to provide all HHS OIG staff with up to date information about HHS grants recipients. The GAP supports OIG's efforts to identify and prevent grants fraud, waste, and abuse.

HHS grants data is currently housed in more than a dozen databases at both the Department and the Operating Division (OPDIV) levels.

The GAP is a collection of analytic products based on a fused database that was designed to connect different grants related data sources. The GAP now puts grants related data at OIG's fingertips.

Scoring/Binning



Data Fusion – Common Identifiers



DUNS Number

Tax ID (EIN)

Organization

Grant Number

PMS Account Number

Text Mining for Grants Analytics



- Grants data comes in lots of different formats
 - Structured data (example: excel files)
 - Semi-structured (example: a report with sections that don't change)
 - Unstructured (example: different types of documents with no set formatting)
- Prior work demonstrated that text mining the A-133 Audit Reports was possible and was meaningful for grants oversight
- In order to convert the A-133 reports to usable data, we tokenized the reports, and then created two predictive models to identify findings pages within the reports

Why Tokenize?



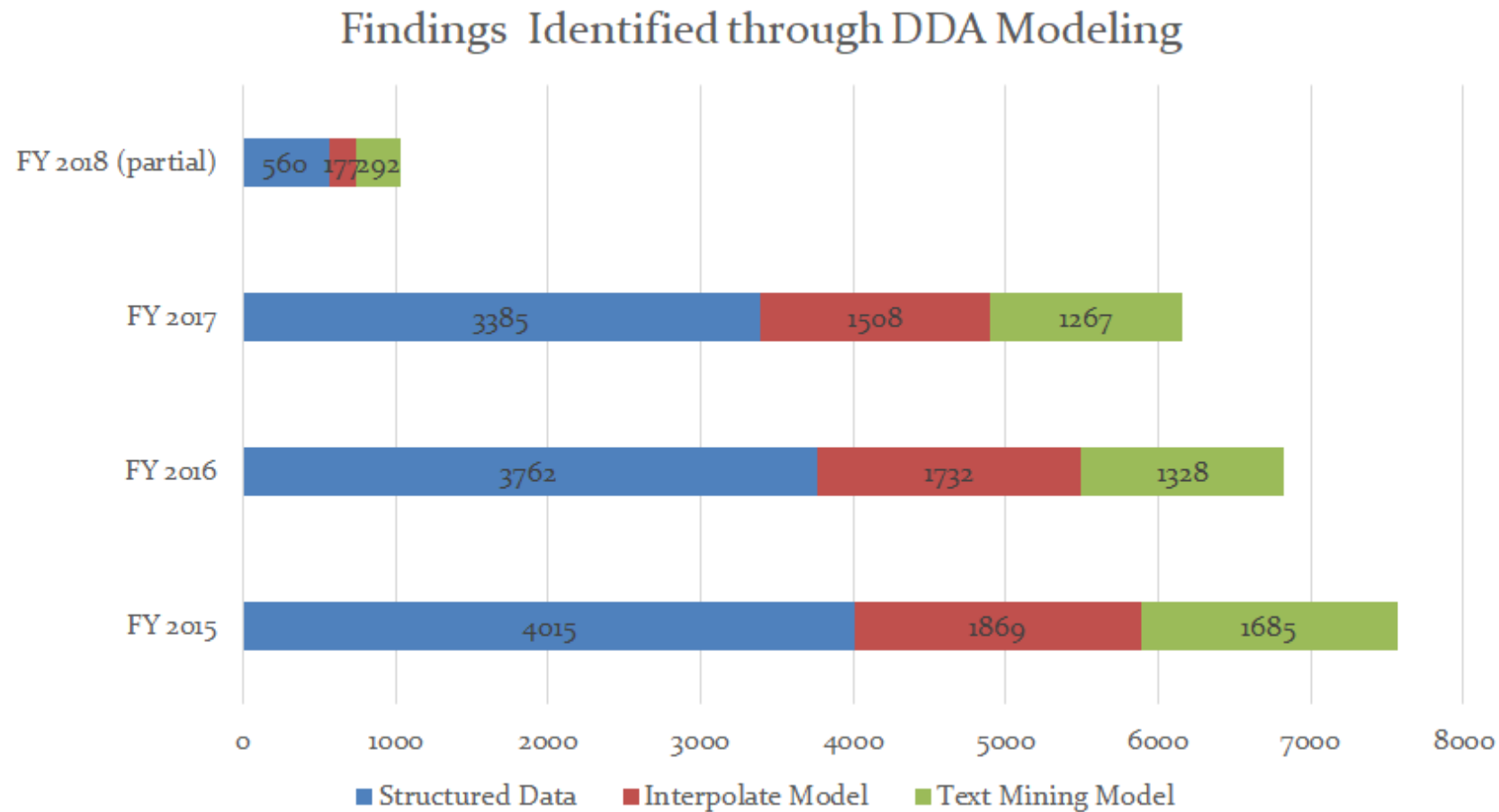
So instead of searching PDFs...

We have a database we can query!



Result 1: Messages				
doc_id	page_id	lines	text	text
2015008100720000	2015008100720018	20	CLIFFORD GLOVER DAY CARE CENTER, INC.	'-17:84'1'62'10:558'47:56'157'63'2015:15'218:58'228:5
2015008101450000	2015008101450040	32	COMMUNITY SERVICES AGENCY OF SOUTH TEXAS, INC.	'-133:122'000:160'2015:21'30:20'300:159'35:170'510:1
2015008100720000	2015008100720023	36	Schedule 1	'042:331'370'045:295'057:96'064:247'066:144'163'067:
122'000:160'2015:21'30:20'300:159'35:170'510:124	93:568:141'a'121,125,154	'accordance:118'agency:13'and:14,155'any:107'are:112'as:163'assistance:144'audit:108'	CLIFFORD GLOVER DAY CARE CENTER	'auditee:161,167'auditor:38,98'auditors:26'awards:70
2015008100720000	2015008100720025	34	Schedule 2	'-24:336'000:154'002:303'013:309'036:302'053:80'054:
2015008100720000	2015008100720026	33	Schedule 2	'-25:299'000:126,132,151'022:120,130,134,137'044:93'0
2015008100720000	2015008100720027	38	Schedule 2	'-26:330'000:287'001:154,313'024:205'026:152'027:156
2015008100720000	2015008100720028	34	Schedule 2	'-27:259'006:86'012:227,229'063:129'064:93'1:124'00:
2015008100720000	2015008100720029	54	Schedule 3	'-28:159'0:00:181'1:22,23,29,30,36,37,43,48,56,69,73,79,
2015008100720000	2015008100720030	52	Schedule 3	'-29:304'0:00:210,214,221,228'1:22,26,38,44,52,76,92,10
2015008100720000	2015008100720031	60	Schedule 3	'-30:237'0:00:65,72,78,133,140'1:22,71,87,88,92,124,132
2015008100720000	2015008100720032	37	Schedule 3	'-31:193'000:15:145,154'1:36,40,44,49,59,72,92,101,105,1
2015008100720000	2015008100720033	42	Schedule 4	'-32:150'075:100'1:18,58,74,106'10:97'100:54,55,102,1
2015008100730000	2015008100730002	17	ERIE COMMUNITY COLLEGE,	'-133:98'2014:23'2015:21'31:20'55:34'56:44'57:74'59:
2015008100730000	2015008100730003	43	Drescher & Maleki LLP	'-1:496'14227:12'2014:82'2015:80'3083:4'31:79'5'8'79
2015008100730000	2015008100730004	42	Options	'-133:390'2:557'2:92'2014:63'2015:51,103'27:127'31:1
2015008100730000	2015008100730005	14	The Other Information, as listed in the table of contents, has not been subjected to the auditing pro	'-31:170'12:63,168'2016:64,169'accordance:50,150'accord
2015008100730000	2015008100730011	39	Table 5--Summary of Expenses (Functional Classifications)--Primary Government	'-9:313'0.3:146'031:75'034:37'054:128'061:42'069:56'
2015008100730000	2015008100730006	48	ERIE COMMUNITY COLLEGE	'-4:612'0.2:296,336'0.5:293,339'0.5:199'1.0:318'1.1:36
2015008001080000	2015008001080019	26	is a deficiency, or a combination of deficiencies, in internal	'-20:272'2016:126,271'29:265'30:270'0:2,5,17,151'about
2015008100730000	2015008100730007	42	Overview of the Financial Statements	'-133:385'5:563'1:35'1.5:448'13.9:427'15.4:434'2:42'
2015008100730000	2015008100730009	45	respectively. These increases are primarily due to the recording of other post-employment benefits,	'-7:1428'0:144'005:377'045:255'056:265'067:342'1:363,
2015008100730000	2015008100730010	43	services provided in return for the operating revenues in carrying out the mission of the College.	'-6:431'0.3:302,343,410'0.4:278'0.8:380'0.9:321'00:19
2015008100730000	2015008100730012	38	Table 6--Summary of Expenses--Primary Government	'-0:0:82,87'0.1:90,99'0.13:137'0.69:124'007:26'054:108
2015008100730000	2015008100730019	89	ERIE COMMUNITY COLLEGE	'000:269,273,304'035:240'037:172'048:399'033:396'054
2015008100730000	2015008100730020	66	ERIE COMMUNITY COLLEGE	'003:153'004:155'010:60,61'028:185'046:133,135'060:7
2015008100730000	2015008100730021	61	ERIE COMMUNITY COLLEGE	'005:378'007:161,164'018:322,324'022:131,133'037:225
2015008100730000	2015008100730022	55	ERIE COMMUNITY COLLEGE	'028:159'054:363'057:94'083:285,288'087:107,110,300,1
2015008100730000	2015008100730023	50	ERIE COMMUNITY COLLEGE	'000:188'001'202,247'023:89'042:126,146'052:127,147'1
2015008100730000	2015008100730024	53	ERIE COMMUNITY COLLEGE	'020:55,58'027:67,70'034:97'040:311'050:298,330'051:
2015008100730000	2015008100730025	57	ERIE COMMUNITY COLLEGE	'000:307'001:306,348'000:375'013'177'104'144'308'0:

Results of A-133 Text Mining Effort



Text Mining/Neural Networks



Finding the right kind of pages

We use a neural network model to “find the finding pages” in the A-133 single audit. This keeps our searches focused.

We want to find things that look like this in the A-133

2013-001 Insufficient cash on hand to cover deferred grant revenue



Then we identify finding numbers

In order to count the number of findings, we count the number of findings numbers on each of the finding pages.

Quiz: how many types of dashes can you find in 35,000 documents?

Future Development

Slice the findings out of the machine readable text

Identify findings that are like the one you might be interested in

Classify different types of findings that an auditor hasn't classified

Look for trends within those classifications

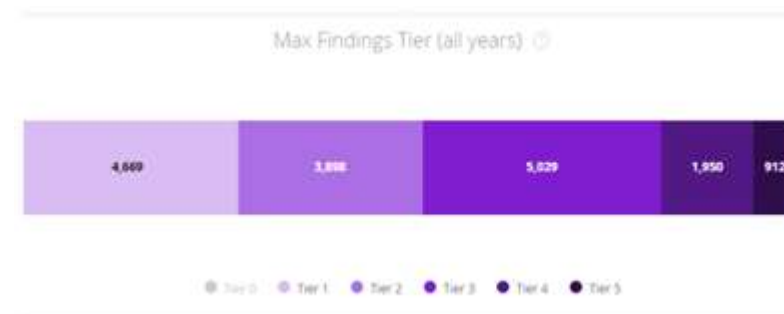
Principal Component Analysis



A-133 single audit components

- The user community identified important aspects of the A-133 single audit reports
- We used a principal component analysis to better understand the relationship between those components
- The outputs of that analysis is used to identify audits with the most interesting mix of components and push them forward for OIG staff

Type C [Cash Management] Findings (Yes / No)	No
Going Concern viability (Yes / No)	No
Questioned Costs (Yes / No)	No
Material Noncompliance (Yes / No)	No
Reportable condition in the financial findings (Yes / No)	No
Reportable condition in the major programs (Yes / No)	Yes
Material weakness in the financial findings (Yes / No)	No
Material weakness in the major programs (Yes / No)	Yes
Significant deficiency in the financial findings (Yes / No)	No
Significant deficiency in the major programs (Yes / No)	Yes
Type of report concerning financial findings	Unmodified Opinion
Type of report concerning major programs	Qualified
Auditee qualified as low risk (Yes / No)	No
HHS Findings	0



Peer Comparison Score – Nearest Neighbors



0.74



Drawdown Amount

Average (this account) = **\$17,592**
Larger drawdowns than **60%** of peers
More variance in drawdowns than **70%** of peers

Drawdown Frequency

Average (this account) = **6.4 days**
More days between drawdowns than **80%** of peers
More variance in frequency than **90%** of peers

Engaging the User Community



Direct Engagements

- Focus Groups
 - Learning about the current business processes of OIG staff – how do things work now
- Persona Development
 - How does the average OIG staff member complete this task?
 - Documenting pain points and challenges
- User Testing
 - How do users work through new features in our tools
- Communities of Practice

Indirect Engagements

- Ad Hoc Requests
 - What kinds of data analysis requests are we working on now?
- Usage Data
 - What features are being used?
- Questions at larger meetings
 - What questions are we getting about features that have been deployed

In Development – Communities of Practice



Within HHS OIG

- Active User Community
 - Feedback
 - New Ideas
 - Test new features
- Share success stories and strategies
- Networking and collaboration where appropriate

Across OIG Community – Informal Collaboration

- Share code and methods
- Share use cases and best practices
- Share challenges and brainstorm solutions



Questions?