

Drinker Biddle

**The Impact of NARA's Email Capstone Policy  
& Other Recent Initiatives on FOIA Access**

Freedom of Information Act Advisory Committee  
Washington, D.C.  
October 19, 2017

Jason R. Baron  
Of Counsel, Drinker Biddle & Reath LLP  
Co-Chair, Information Governance Initiative

© Jason R. Baron 2017

## Outline of Talk

- Some preliminary observations about the e-recordkeeping landscape in 2017
- The US Archivist's 2012 Directive: A Framework for A New Age of Archives in the U.S.
- The Capstone Email Policy
- Initial returns on agencies meeting M-12-18 deadlines
- FOIA Access to Digital Archives: Barriers & Solutions
- Recommendations

# Drinker Biddle



## Yesterday's Archival Collections



DrinkerBiddle

# Tomorrow: Exponentially Growing Amounts of Digital Information



## Growth in Presidential and Federal Records in Electronic Form

- William Clinton White House 32 million emails (presidential and federal records)
- George W. Bush White House 200+ million emails
- Barack Obama White House ~ 300+ million emails received in Jan. 2017 + records on social media
- And after 2019: billions of emails and all forms of electronic records from hundreds of Executive Branch agencies being accessioned into the National Archives
- **HOW MANY OF THESE RECORDS WILL BE ACCESSIBLE TO THE PUBLIC AND HOW SOON?**

## DrinkerBiddle

Somewhere in a vast collection of digital objects stored in archives are important records relevant to FOIA requests, e-discovery demands, investigations, audits, as well as to historians and civil society at large. How can we efficiently access them?



DrinkerBiddle

# A New Era of Government

**“[P]roper records management is the backbone of open Government.”**

**President Obama’s Memorandum dated November 28, 2011**

**re “Managing Government Records”**

**<http://www.whitehouse.gov/the-press-office/2011/11/28/presidential-memorandum-managing-government-records>**



THE WHITE HOUSE

Office of the Press Secretary

For Immediate Release

November 28, 2011

November 28, 2011

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

SUBJECT: Managing Government Records

Section 1. Purpose. This memorandum begins an executive branch-wide effort to reform records management policies and practices. Improving records management will improve performance and promote openness and accountability by better documenting agency actions and decisions. Records transferred to the National Archives and Records Administration (NARA) provide the prism through which future generations will understand and learn from our actions and decisions. Modernized records management will also help executive departments and agencies (agencies) minimize costs and operate more efficiently. Improved records management thus builds on Executive Order 13589 of November 9, 2011 (Promoting Efficient Spending), which directed agencies to reduce spending and focus on mission-critical functions.

When records are well-managed, agencies can use them to assess the impact of programs, to reduce redundant efforts, to save money, and to share knowledge within and across their



EXECUTIVE OFFICE OF THE PRESIDENT  
OFFICE OF MANAGEMENT AND BUDGET  
WASHINGTON, D.C. 20503

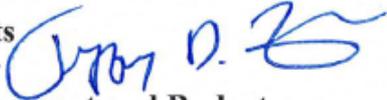


NATIONAL ARCHIVES AND RECORDS ADMINISTRATION  
WASHINGTON, D.C. 20408

August 24, 2012

M-12-18

**MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND  
AGENCIES AND INDEPENDENT AGENCIES**

**FROM:** Jeffrey D. Zients  
Acting Director   
Office of Management and Budget

David S. Ferriero   
Archivist of the United States  
National Archives and Records Administration

**SUBJECT: Managing Government Records Directive**

On November 28, 2011, President Obama signed the [Presidential Memorandum – Managing Government Records](#). This memorandum marked the beginning of an Executive Branch-wide effort to reform records management policies and practices and to develop a 21st-century framework for the management of Government records. The expected benefits of this effort include:

# Archivist/OMB Directive

- M-12-18, Managing Government Records Directive, dated 8/24/12:
  - 1.1 By 2019, Federal agencies will manage all permanent records in an electronic format.
  - 1.2 By 2016, Federal agencies will manage both permanent *and* temporary email records in an accessible electronic format.

<http://www.whitehouse.gov/sites/default/files/omb/memoranda/2012/m-12-18.pdf>

---

The logo for Drinker Biddle, featuring the company name in a white serif font on a blue rectangular background. The background of the slide is a light gray grid pattern.

Drinker Biddle

## From the 2016 Federal Agency Records Management Annual Report:

“Our analysis of the SAORM reporting data shows that M-12-18 is changing Federal records management from paper-intensive analog-based methods to a digital government as intended.”

Source: <https://www.archives.gov/files/records-mgmt/resources/2016-federal-agency-records-management-annual-report.pdf>

## From the 2016 Fed Agency Records Mgmt Annual Report:

- • Seventy-nine (79) percent of agencies report meeting the OMB/NARA Managing Government Record Directive (M-12-18) target to manage all email in electronic format by December 31, 2016. The remaining 21% report progress.
- Ninety-eight (98) percent of agencies are optimistic about meeting the goal to manage all permanent electronic records in electronic format by December 31, 2019.
- • Printing and filing as a method of capturing email continues to decline as indicated (75% of agencies in 2013 to 46% in 2016), with email archiving trending in opposite direction (27% of agencies in 2013 to 80% in 2016) (percentages sum to over 100% as many agencies use multiple approaches).
- Source: <https://www.archives.gov/files/records-mgmt/resources/2016-federal-agency-records-management-annual-report.pdf>

## NARA Draft Strategic Plan 2018-2022

- **“By FY 2020, NARA will have policies and processes in place to support Federal agencies’ transition to fully electronic recordkeeping.** We added this new objective under our Strategic Goal *Connect with Customers* based on comments from our customer Federal agencies who asked us to make a commitment to assist them in transitioning to a fully electronic environment.”
  - Source: <https://www.archives.gov/about/plans-reports/strategic-plan/draft-strategic-plan> (bold in original)
-

## Failed\* E-mail Recordkeeping Paradigms

- Print to paper
- Backup tapes
- Idiosyncratic e-recordkeeping in user folders
- DoD 5015.2\*

\*not quite “failed,” but under-utilized

## Capstone Approach

- Addresses challenges in managing email
- Preserve permanently valuable email and dispose of temporary email

# Capstone Approach

Categorize and schedule email based on the position and/or work of the email account owner

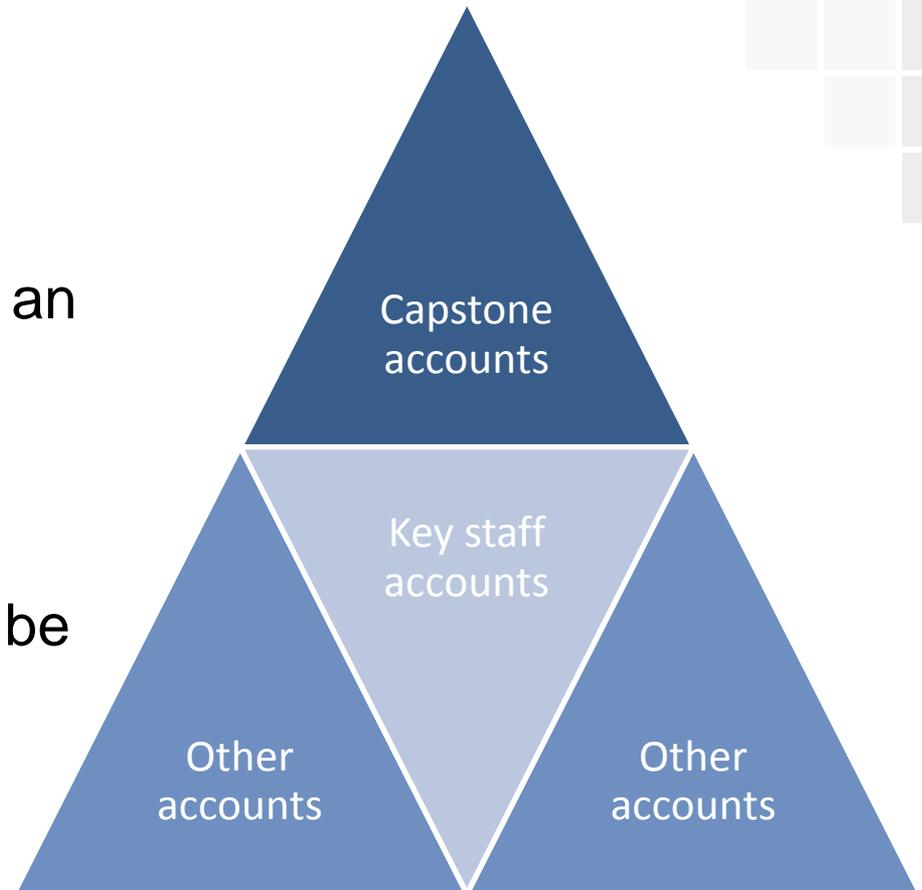
Schedule all of the email belonging to selected officials as permanent

Schedule all other email accounts as temporary to be preserved for a set period of time based on agency needs

# Capstone Officials

Capstone officials may include:

- Officials at or near the top of an agency or an organizational subcomponent
- Key staff members that may be in positions that create or receive presumptively permanent email records



## GRS 6.1—Email managed under Capstone

### GENERAL RECORDS SCHEDULE 6.1: EMAIL MANAGED UNDER A CAPSTONE APPROACH

Capstone Accounts: Permanent

Non-Capstone Accounts: Email retained for 7 years (for substantive records)

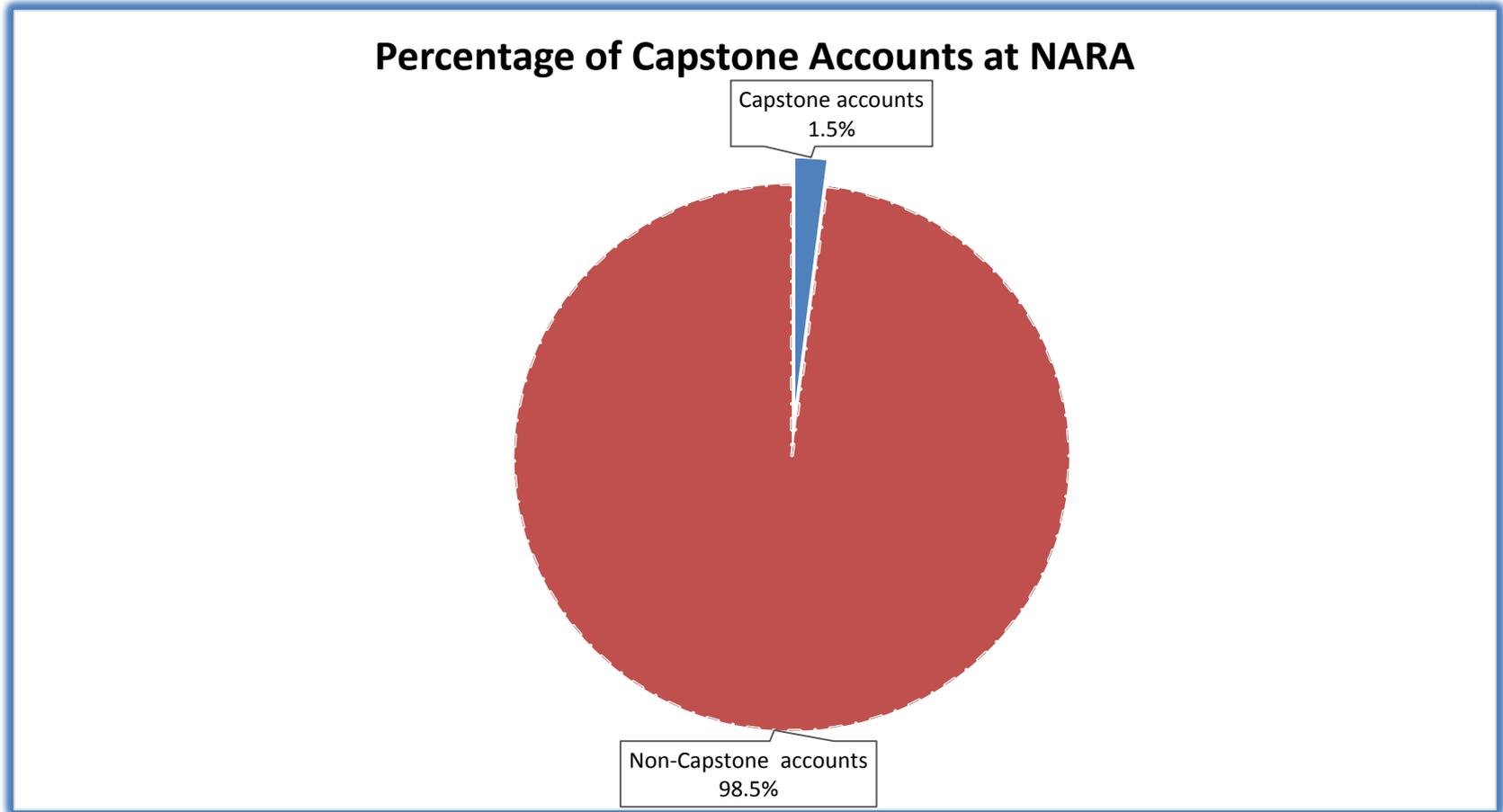
Non-Capstone Accounts: Email retained for 3 years (administrative records)

Agencies must submit a form listing names of Capstone accounts enterprise-wide before they may take advantage of the GRS 6.1.

**From 6.1: “The Capstone Approach and Capstone GRS Agencies are expected to apply documented selection criteria to cull the email of Capstone officials (permanent accounts) to the greatest extent possible before transfer to NARA. Culling refers to the removal – or otherwise excluding from capture – of nonrecord, personal, or transitory messages and attachments.”**

See <https://www.archives.gov/files/records-mgmt/grs/grs06-1.pdf>

## One example of % Capstone Accounts at an Agency



## Capstone agencies / number of Capstone accounts

- [ADMINISTRATION FOR CHILDREN AND FAMILIES](#) 46
- · [Centers for Disease Control and Prevention](#) 35
- · [Centers for Medicare & Medicaid Services](#)
- · [Consumer Financial Protection Bureau](#)
- · [Department of Education](#) 45
- · [Department of Energy, National Nuclear Security Administration](#)
- · [Department of Health and Human Services, National Institutes of Health](#)
- · [Department of Labor, Occupational Safety and Health Administration](#)
- · [Department of the Army \(5 components\)](#)
- · [Department of the Navy, Chief of Naval Operations \(CNO, OPNAV\)](#)
- · [Department of the Navy, Secretary of the Navy \(SECNAV\)](#) 208
- · [Department of the Treasury](#) 244
- · [Environmental Protection Agency](#) 110
- · [Export-Import Bank of the United States](#)
- · [Federal Energy Regulatory Commission](#)
- · [General Services Administration](#) 135

## More Capstone agencies

- · [Library of Congress](#)
- · [National Archives and Records Administration](#) 45
- .. [National Geospatial-Intelligence Agency](#)
- · [National Security Agency/Central Security Service](#) 394
- · [Nuclear Regulatory Commission](#)
- · [Pension Benefit Guaranty Corporation](#)
- .. [Securities and Exchange Commission](#)
- · [Selective Service System](#)
- · [Social Security Administration](#) 363
- · [Substance Abuse and Mental Health Services](#)
- · [U.S. Department of Energy/Energy Information Administration \(EIA\)](#)
- · [U.S. OFFICE OF SPECIAL COUNSEL](#)
- · [U.S. Office of Personnel Management](#)
- · US Dept of Agriculture (11 components)
- SOURCE: <http://usnationalarchives.github.io/capstone-grs/>

## NARA's Capstone Success Criteria: Systems

- **Systems:** Agencies must have systems in place that can produce, manage, and preserve email records in an acceptable electronic format until disposition can be executed. Additionally, systems must support the implementation of agency policies and provide access to email records throughout their lifecycle.
- **What Success Looks Like:** Your agency's systems and business processes support the management of email records in accordance with all applicable requirements including the manual or **automatic execution of their disposition whether using a Capstone-based or content-based record schedule.**

Source: <https://www.archives.gov/files/records-mgmt/email-management/2016-email-mgmt-success-criteria.pdf>

## NARA's Capstone Success Criteria: Access

- Access: Email records must remain usable and retrievable throughout their lifecycle. Access supports an agency's ability to carry out its business functions. Access should address internal agency needs and accommodate responses to requests for information.
- What Success Looks Like: Your agency's email records are maintained in a system that preserves their content, context and structure, protects against their unauthorized loss or destruction, and ensures that they remain discoverable, retrievable, and usable for the period specified in their retention schedule.

Source: <https://www.archives.gov/files/records-mgmt/email-management/2016-email-mgmt-success-criteria.pdf>

## Alternative Success Criteria

- Did agencies *really* adopt Capstone or just implement forms of “save everything” policies?
- Will transfer guidance for 2019 compel agencies to provide public-use versions of records without PII content?
- How do agencies cope with exponentially growing amounts of email records in email archives? Are they employing more robust search technologies?
- Will NARA recommend seamless integration of permanent emails with other related permanent records after 2019?
- How are agencies dealing with FOIA now for email?

DrinkerBiddle

**The Coming Age of Dark Archives (i.e., the inability to provide access unless we have smart ways of extracting signal from noise, including use of privacy filters)**



DrinkerBiddle

## Windows for Citizen Access to Dark Archives....



**Open Data  
Policies**

**Freedom  
OF  
Information  
— ACT —**



## NARA 1601 on Screening Records (2002)

- Screen records if there is a reasonable chance that they may contain information about a living individual that reveals details of a highly personal nature, which if released would constitute a clearly unwarranted invasion of privacy. Withhold such information from files before disclosure if it has not been officially released previously or if it relates to individuals less than 75 years old or events that occurred less than 75 years before the date of screening.

Source: <https://www.archives.gov/foia/directives/nara1601.pdf>

---

DrinkerBiddle

Anticipating the need to filter sensitive content of all types



## Categories of Sensitive Info in Records (non-exhaustive)

- Personally identifiable information (PII)
  - Names as metadata fields; social security numbers; telephone numbers; driver's license information; taxpayer information; bank or financial account information; credit card numbers; vehicle registrations; dates of birth, height and weight characteristics; asset information
- Personal health information (PHI)
- Arrest records
- National security & law enforcement investigations
- Intellectual property
- Material covered by a legal privilege
- FOIA exempt material (in U.S., nine categories)

## Types of sensitive content

- Content susceptible to deletion through the use of “regular expressions”
- Textual content
- Nontextual content (images, fingerprints, biometric data, geospatial imaging)

## Leveraging the Archivist's Directive to obtain more meaningful access to public records

- The 2016 and 2019 directives mean electronic mail and permanent electronic records of all forms will be preserved in searchable formats
- Adoption of Capstone policies mean that a repository of emails and attachments from senior officials can easily be made subject to a “reasonable search” under FOIA
  - With all other substantive emails preserved for at least 7 years
- A more complete historical record of government will be available to future generations, *if* barriers to access can be overcome.

# An information retrieval task: searching the Haystack....



DrinkerBiddle

to find relevant needles...



## Drinker Biddle

But not just one or a few  
needles...

Lawyers need to  
find ALL the  
relevant needles  
(100% recall)

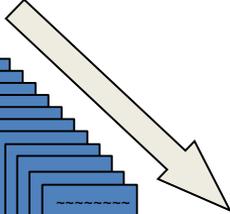
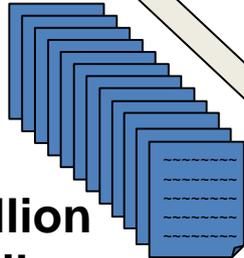


And they would  
like to find JUST  
the wheat, not  
the chaff (100%  
precision)



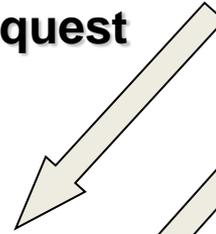
**Clinton  
White House**

**20 million  
emails**



**National Archives**

**search  
request**

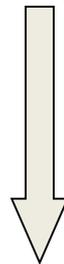


**Tobacco  
Policy**

**100,000  
relevant  
docs, with  
20,000  
privileged**



**hired 25  
persons**

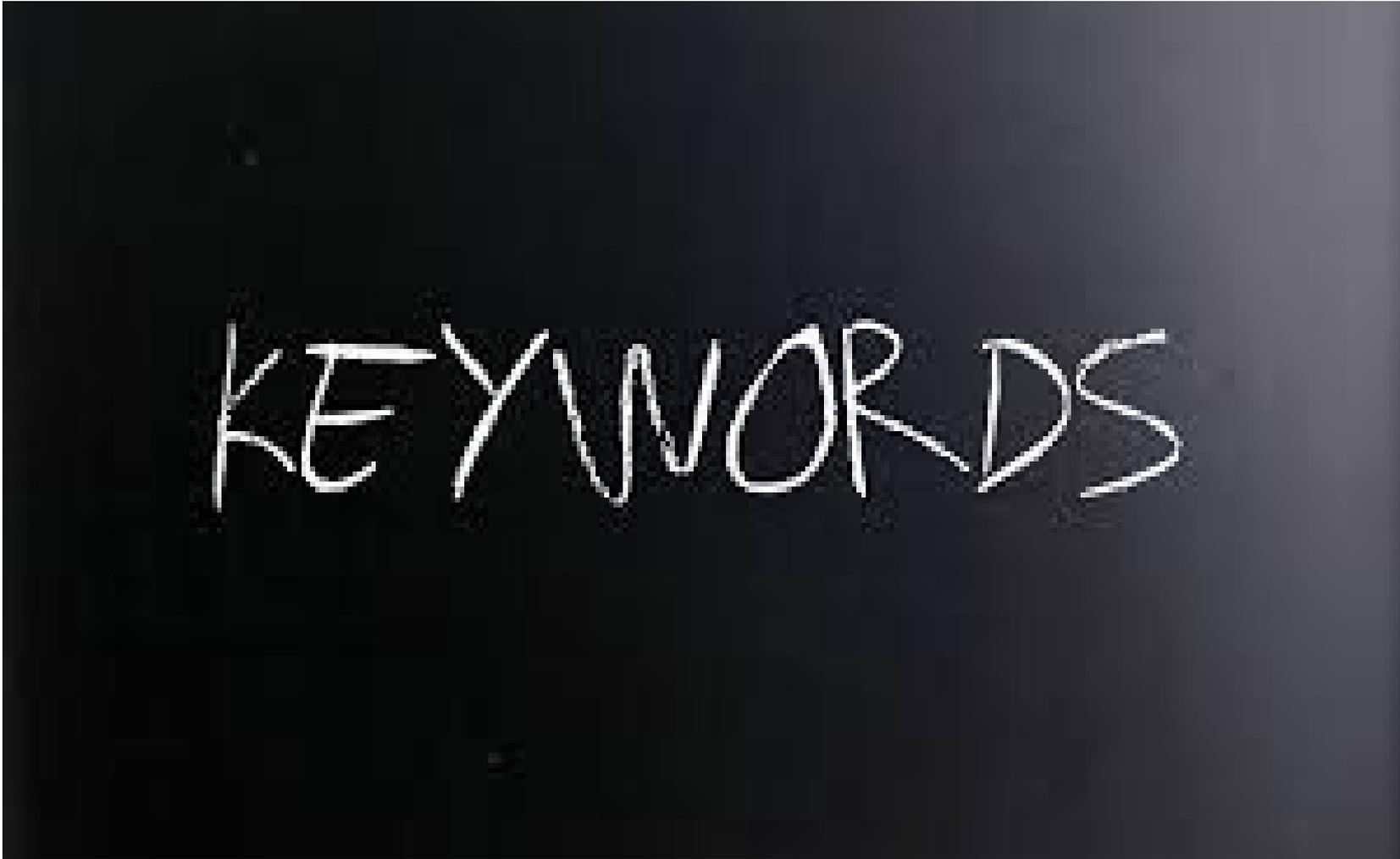


**for 6 months ...**



**200,000 docs with  
“hits”**

A flawed, inefficient approach to searching ESI.....

A photograph of a chalkboard with the word "KEYWORDS" written in white chalk. The letters are slightly irregular and spaced out, typical of handwriting. The background is a dark, textured grey.

KEYWORDS

# In Search of A Better Way for Lawyers To Search ESI

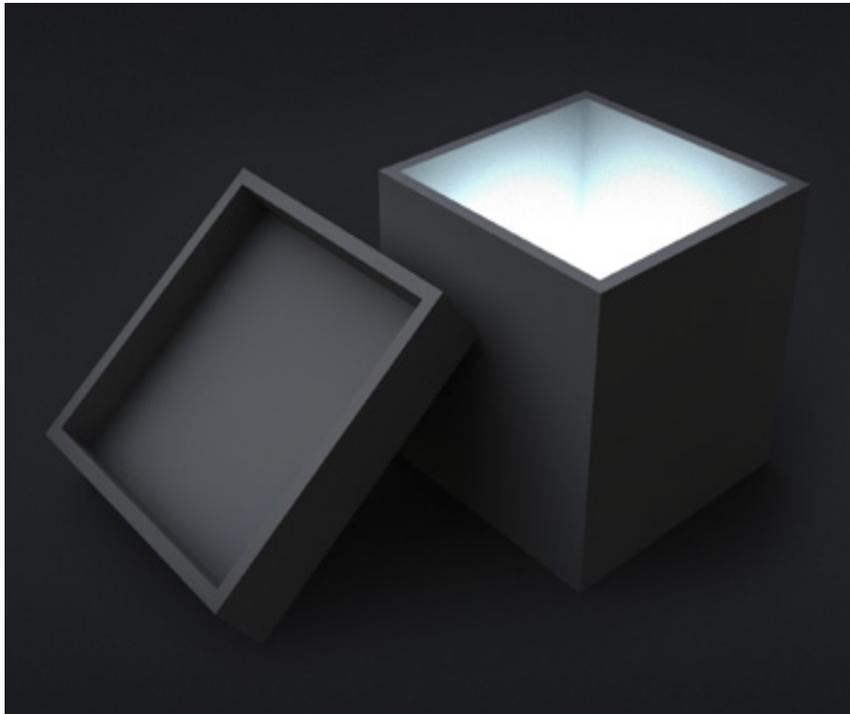


A word cloud centered around the theme of predictive coding. The words are arranged in a roughly circular pattern, with varying font sizes and colors. The largest words are 'Predictive', 'Coding', and 'Review'. Other prominent words include 'Automated', 'Intelligent', 'Assisted', 'Based', 'Computer-Assisted', 'Technology', 'Adaptive', 'Meaning', 'Classification', 'Automatic', 'Transparency', 'Privilege', 'Document', 'Tagging', 'Advanced', 'Prioritization', 'Ranking', 'Presumptive', and 'AutoSuggest'.

Automated Intelligent  
Coding  
Assisted  
Predictive  
Based Computer-Assisted  
Technology Adaptive  
Meaning  
Classification Automatic  
Transparency Privilege Document  
Tagging Advanced  
Prioritization Ranking  
Presumptive  
AutoSuggest

DrinkerBiddle

Black boxes: how do algorithms separate wheat and chaff?

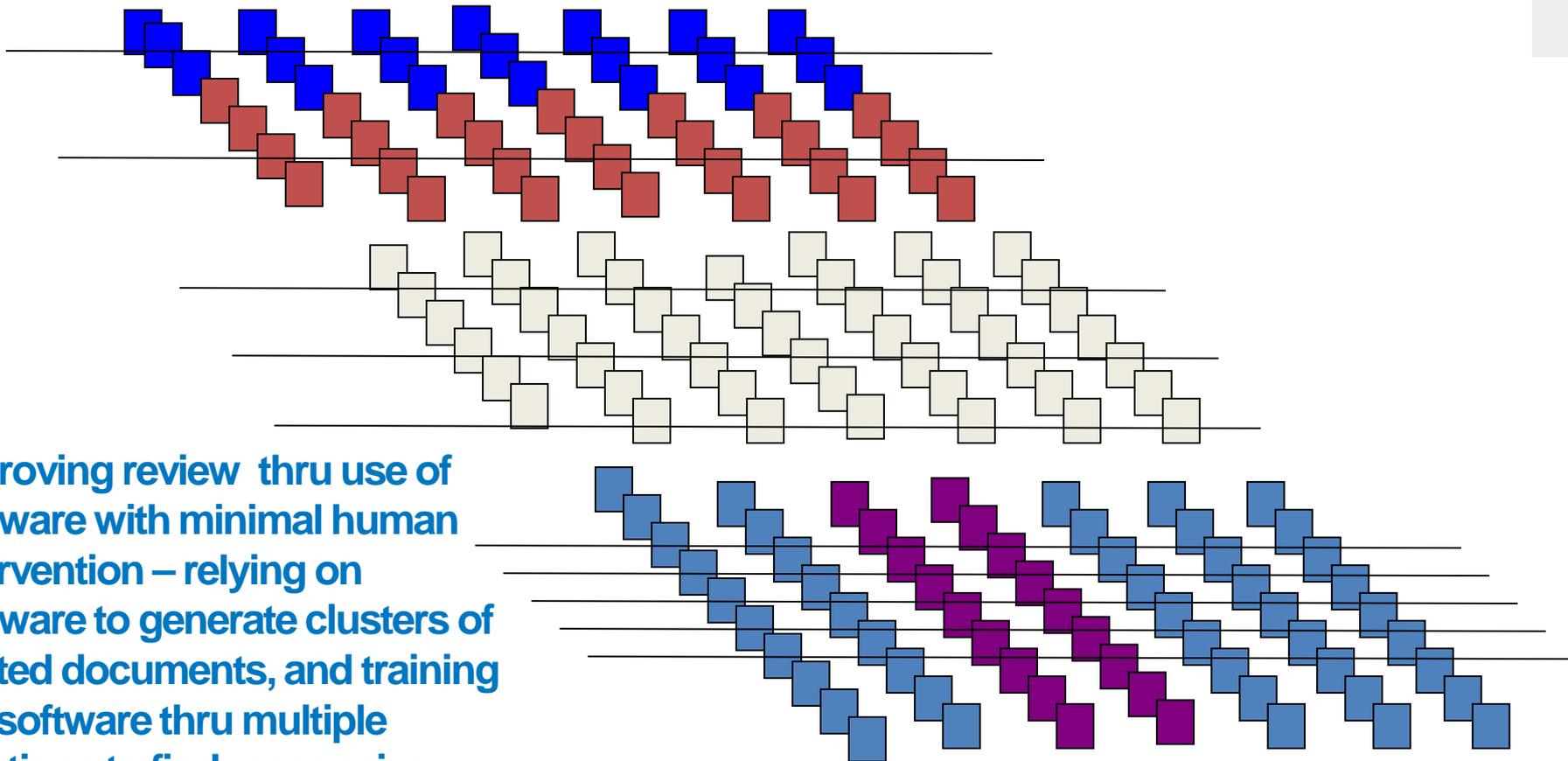


## Defining “predictive coding” or “TAR”

- A process for prioritizing or coding a collection of electronic documents using a computerized system that harnesses human judgments of one or more subject matter experts on a smaller set of documents and then extrapolates those judgments to the remaining document population.
- Also referred to as “supervised or active machine learning,” “computer-assisted review” or “technology-assisted review”

*Source: Adapted from Grossman-Cormack Glossary of Technology Assisted Review, v. 1.0 (Oct 2012)*

# E-discovery Strategies: “Predictive Coding”

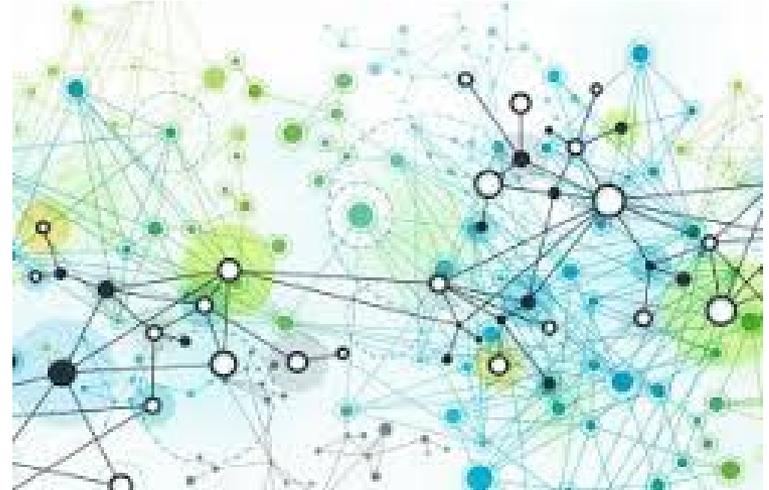
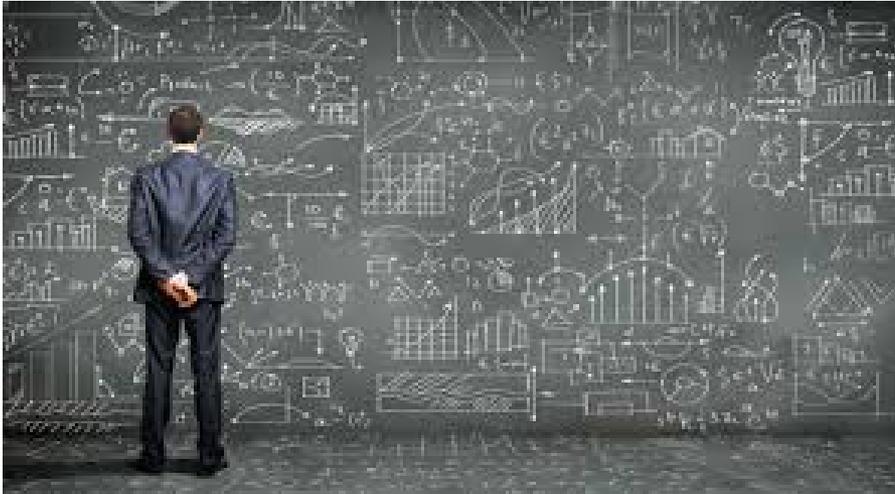


Improving review thru use of software with minimal human intervention – relying on software to generate clusters of related documents, and training the software thru multiple iterations to find responsive records

## **Judicial endorsement of predictive analytics in document review by Judge Peck in *da Silva Moore v. Publicis Groupe* (SDNY Feb. 24, 2012)**

This opinion appears to be the first in which a Court has approved of the use of computer-assisted review. . . . What the Bar should take away from this Opinion is that computer-assisted review is an available tool and should be seriously considered for use in large-data-volume cases where it may save the producing party (or both parties) significant amounts of legal fees in document review. Counsel no longer have to worry about being the ‘first’ or ‘guinea pig’ for judicial acceptance of computer-assisted review . . . Computer-assisted review can now be considered judicially-approved for use in appropriate cases.

# Faith in Analytics



DrinkerBiddle

## Beyond Capstone: Automated Categorization of Records into Records Schedule Categories



## Ensuring a complete record of governmental activities

- Proposition 1: encouraging the use of capture and filter technologies to ensure that there will be a complete historical record, and access to it
- Proposition 2: encouraging auto-categorization of records content to assist in meeting RM and access requirements
- Proposition 3: e-government actors should be promoting the use of advanced search techniques to find and open public records as soon as possible, taking into account sensitive content.

## Research agenda

- Signal to noise ratio: analyzing Capstone's success in capturing (and only capturing) e-mail records from senior officials that are of permanent value;
- Appraising existing archival strategies for filtering sensitive content in large e-mail repositories; and
- Evaluating how well total capture of email has increased the right of citizen access to these records, through informal and formal means.

DrinkerBiddle

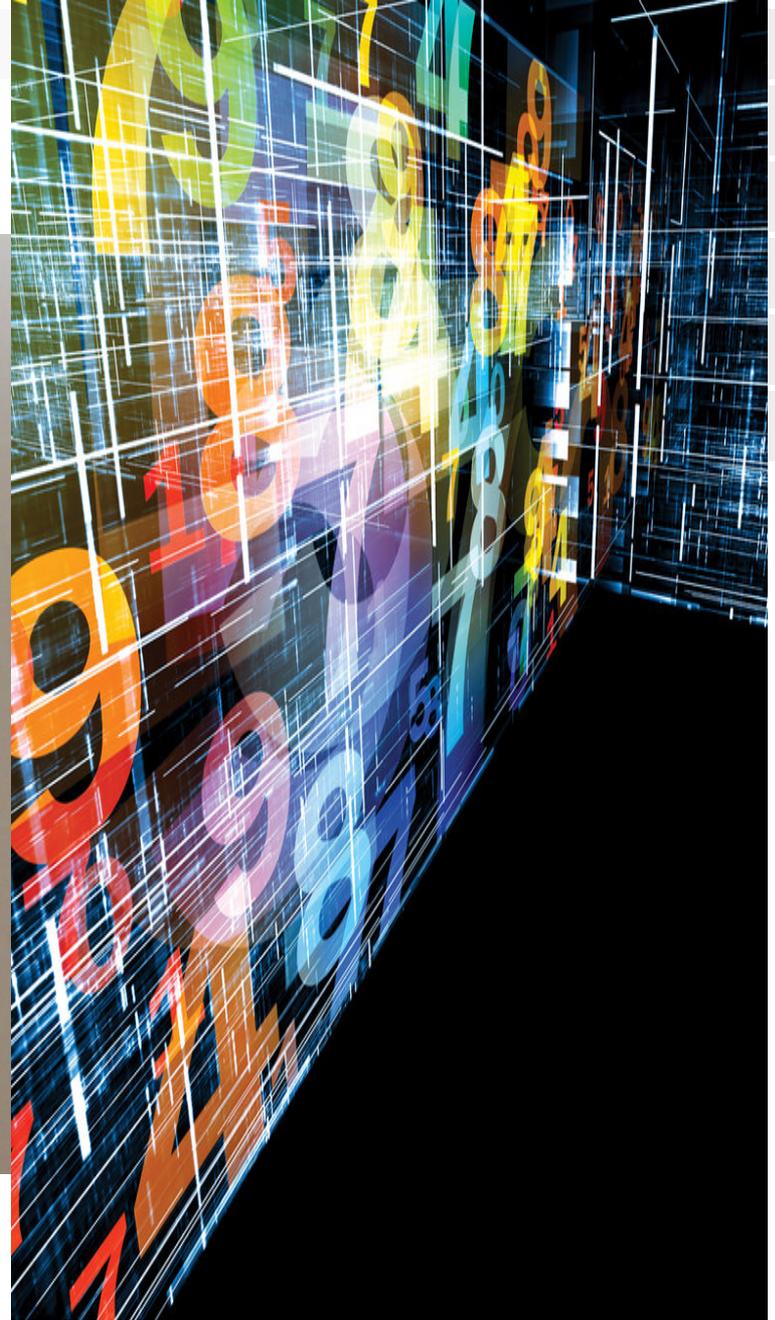
## Recommendations



## Recommendations for this Committee

- FOIA staff need to understand the profound impact the adoption of NARA's Capstone policy will have on their agency
- FOIA staff should be urged to work with General Counsel's offices, records managers, and IT staff in understanding agency capabilities to perform advanced forms of searches of Capstone repositories
- NARA and the scientific community should be looking at R&D on autocategorization, advanced search, and filtering technologies, so as to enable greater access to "dark" repositories and archives of digital objects due to the presence of sensitive/exempt content.

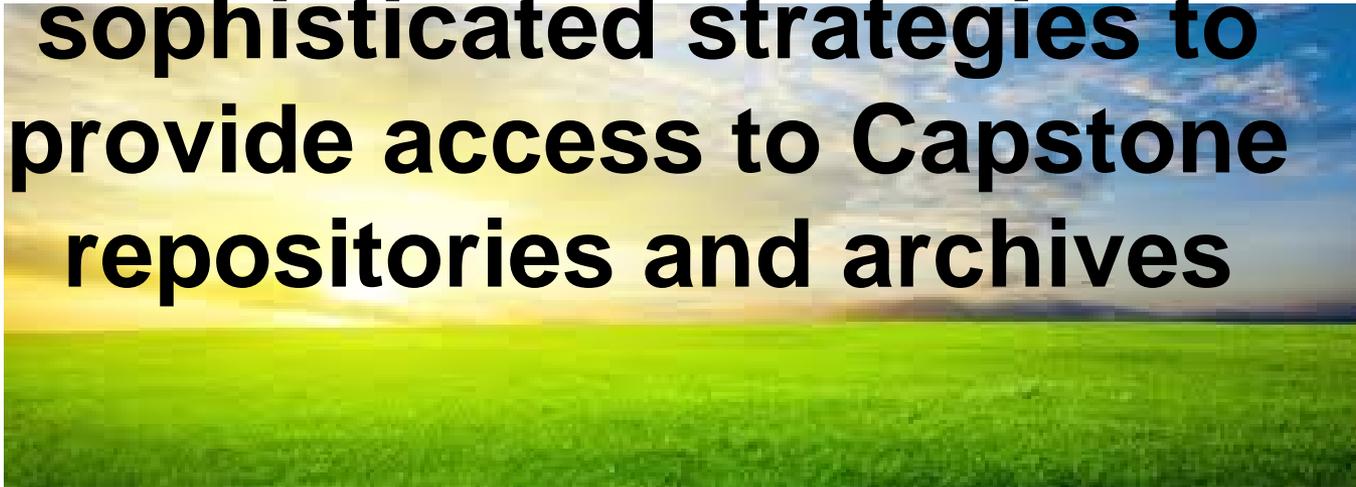
# DrinkerBiddle



Sunsetting of older employees in the current workforce...



**... means a green field for  
millennial FOIA officers, records  
managers, and archivists to take  
up this Call to Arms & insist on  
using more technologically  
sophisticated strategies to  
provide access to Capstone  
repositories and archives**



DrinkerBiddle

To Advance the Cause of  
Promoting FOIA Access, Lawyers  
& RM & IT & FOIA Staff Need To  
Cross Intellectual Boundaries

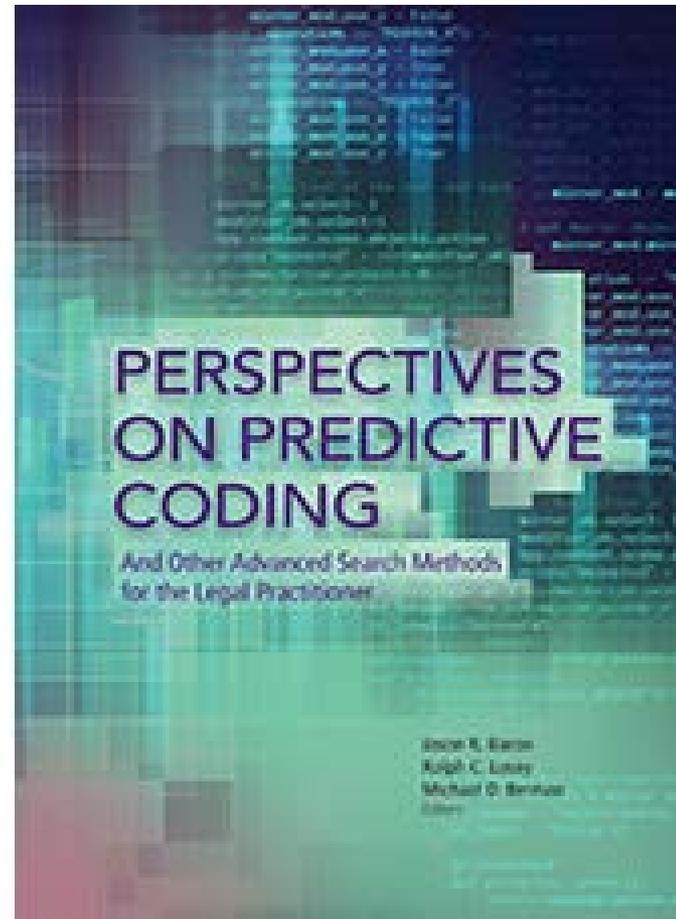


## DrinkerBiddle

Culture change is possible (even for government agencies adopting advanced searching and filtering approaches in meeting their FOIA obligations)

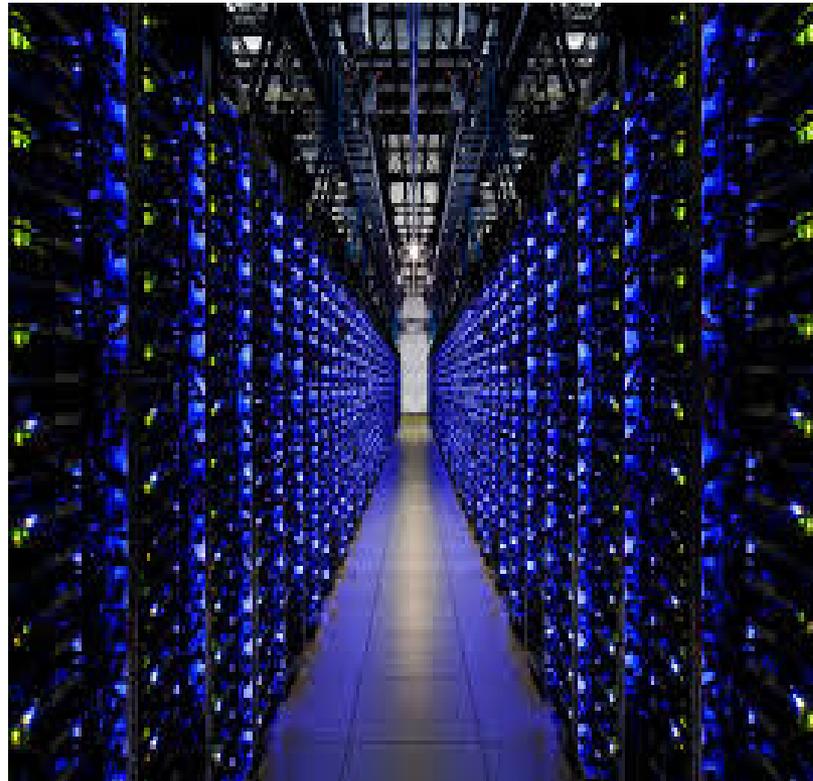


## Perspectives on Predictive Coding



Jason R. Baron  
Drinker Biddle LLP  
[jason.baron@dbr.com](mailto:jason.baron@dbr.com)  
202.230.5196  
*Twitter:* @jasonrbaron1

Information Governance  
Initiative  
[www.iginitiative.com](http://www.iginitiative.com)



© 2017 The views expressed are the author's alone and not any firm or institution with which he is affiliated.