



# **Lawyering in the Information Age:**

**Leveraging Information to Achieve  
Legal Outcomes**

**2015 Annual Government  
Civil Practice Conference**

**October 16, 2015**

**Moab, UT**

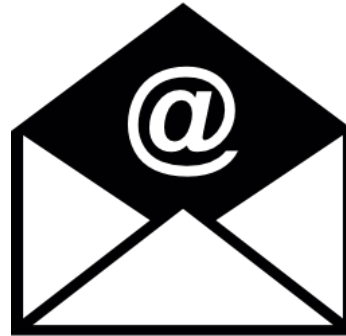
# Information is Everything

- A lawyer's **most essential resource** is information.
- It is also her **most fundamental product**.
- Understanding the client's position or circumstance is **the foundation of our work**.
- It provides **the basis of all of our counsel** regardless of the legal context.

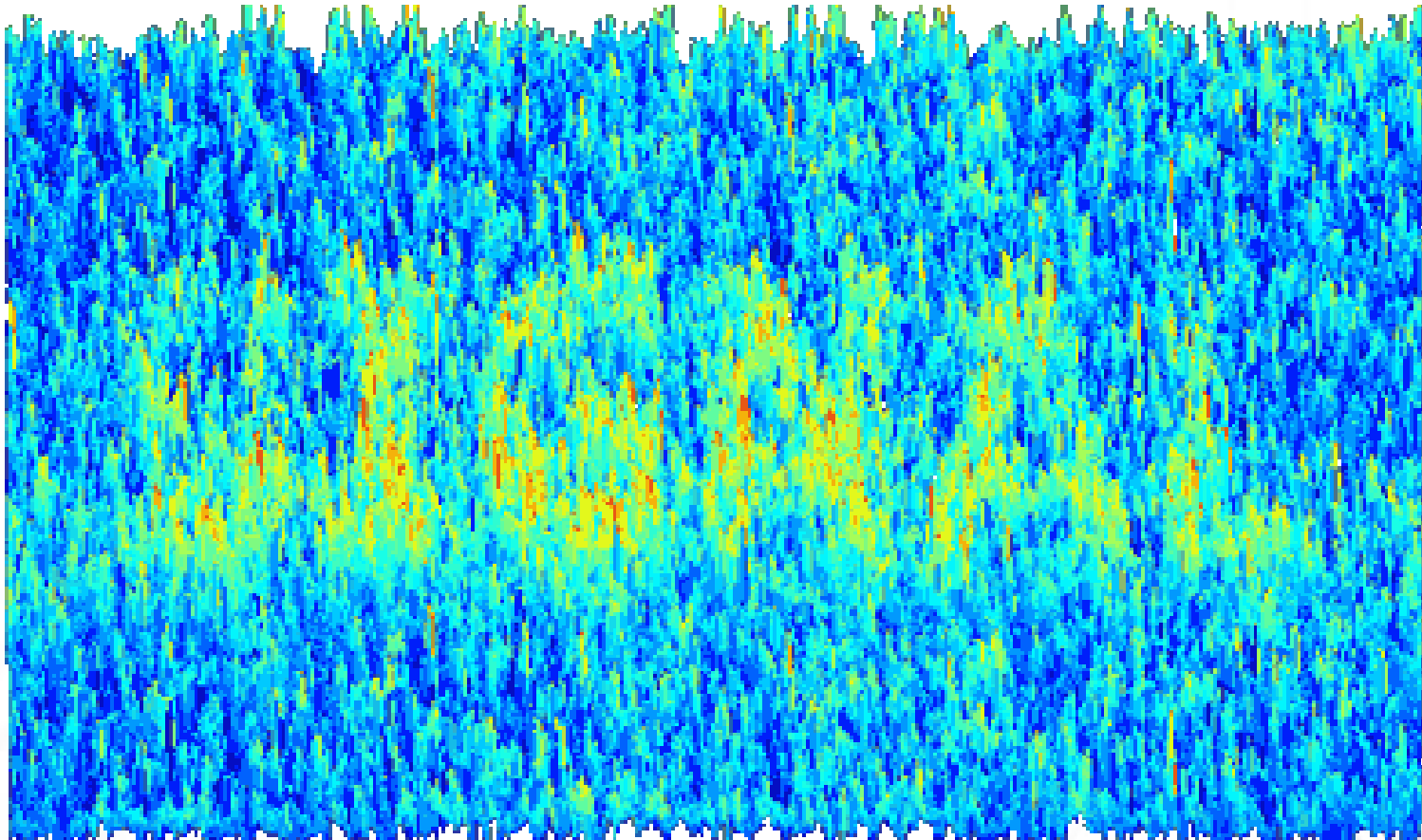


# Information is Overwhelmingly Electronic

- Body cameras
- E-mail
- Electronic documents
- Chat
- Text
- Facebook, Twitter, Instagram



# **Our Challenge: Finding the Signal in the Noise**



**Don't be a Dinosaur!**



# Learn to be Facile with Information

- There are a **finite number of sources**, and only some of those are commonly critical.
- **Learn about these sources**, how to get data out of them, and the relative burden of doing so.
- **Data isn't as complicated as you think**. Use your formidable reasoning skills to **develop a strategy** and **execute it defensibly**.



# Basics: Duty to Preserve

**If you don't get this right, little else matters**

- Requires a party to **identify, locate, and maintain** information and tangible evidence that is relevant to specific and identifiable litigation
- Arises when litigation is “reasonably anticipated,” “reasonably foreseeable,” or “**reasonably likely**”
- Failure to preserve leads to **spoliation**, which can lead to sanctions
- Failure to preserve is the **most common cause of all discovery sanctions.**

# Strategically Identify the Information You Need

- Focus on the Elements of the **Claims and Defenses**
  - This is also the scope of discovery
- Understand the **subject matter** of the case
- Identify the **people** who relate to the subject matter
- Identify the **data sources** they have access to
- Focus on preserving the **most important data** and the data **most at risk** of loss or alteration



# Not All Data is Equal: Set up an Iterative Approach

- Consider **grouping custodians** into tiers
- Focus on **easily accessible data sources first**
- Prove up **relative benefit and burden**
- **Don't argue over the unknown**
- You can usually **agree on a few** custodians, sources and search parameters.
- **Let the data argue for itself**

# Use Proportionality

- Utah Rule 26
  - (b)(2) Discovery and discovery requests are proportional if:
    - (A) the discovery is **reasonable**, considering the needs of the case, the amount in controversy, the complexity of the case, the parties' resources, the importance of the issues, and the importance of the discovery in resolving the issues;
    - (B) the likely **benefits of the proposed discovery outweigh the burden** or expense;
    - (C) the discovery is consistent with the overall case management and will further the **just, speedy and inexpensive determination** of the case;
    - (D) the discovery is **not unreasonably cumulative or duplicative**;
    - (E) the information **cannot be obtained from another source** that is more convenient, less burdensome or less expensive; and
    - (F) the party seeking discovery **has not had sufficient opportunity** to obtain the information by discovery or otherwise, taking into account the parties' relative access to the information.
  - (b)(3) The party seeking discovery always has the burden of showing proportionality and relevance.



# Prove Up Proportionality



- This is a **balancing test**
- You have to have evidence on **BOTH** sides of the scale (benefit and burden)
- **Statistical sampling** is the best way to prove the benefit
- Obtain a **base responsiveness rate**
- **Measure every search parameter against this rate**
- Note: **THEY** have to prove proportionality. So make them,

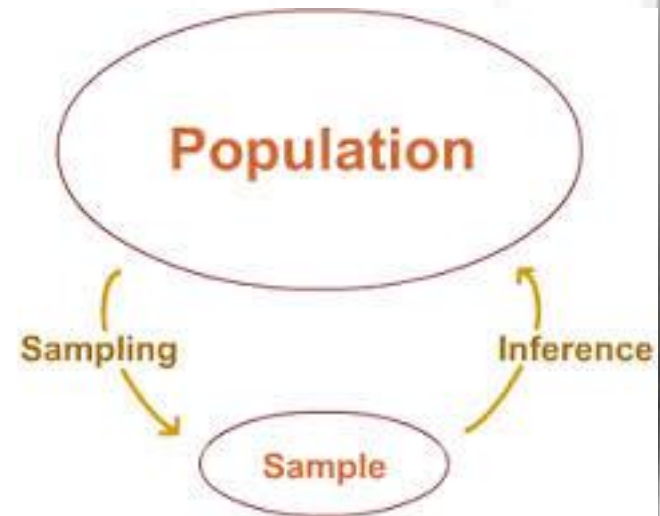
# Rebalancing Asymmetric Litigation

- **Build a fortress** around your own ediscovery practices
- Then **go on the offensive**
- **Attack their preservation**
- **Go after** email accounts, personal computers and devices, social media, pictures
- This is especially **useful in class action cases**



# Statistical Sampling in Plain English

- The **population** is the big batch of data about which you're trying to draw conclusions
- A **Sample** is a portion of the population
- In order to be valid, the sample has to be **representative**, it has to have the characteristics of the population
- The sample also has to be the **right size**
- If it is representative and the correct size, you can prove that the larger population has the **same characteristics as the sample.**
- This is also how predictive coding works



# Getting to Information Faster and More Accurately Leads to Better Lawyering





# **Analytics is the Key to Faster, More Accurate Information**



# Analytics in the Marketplace





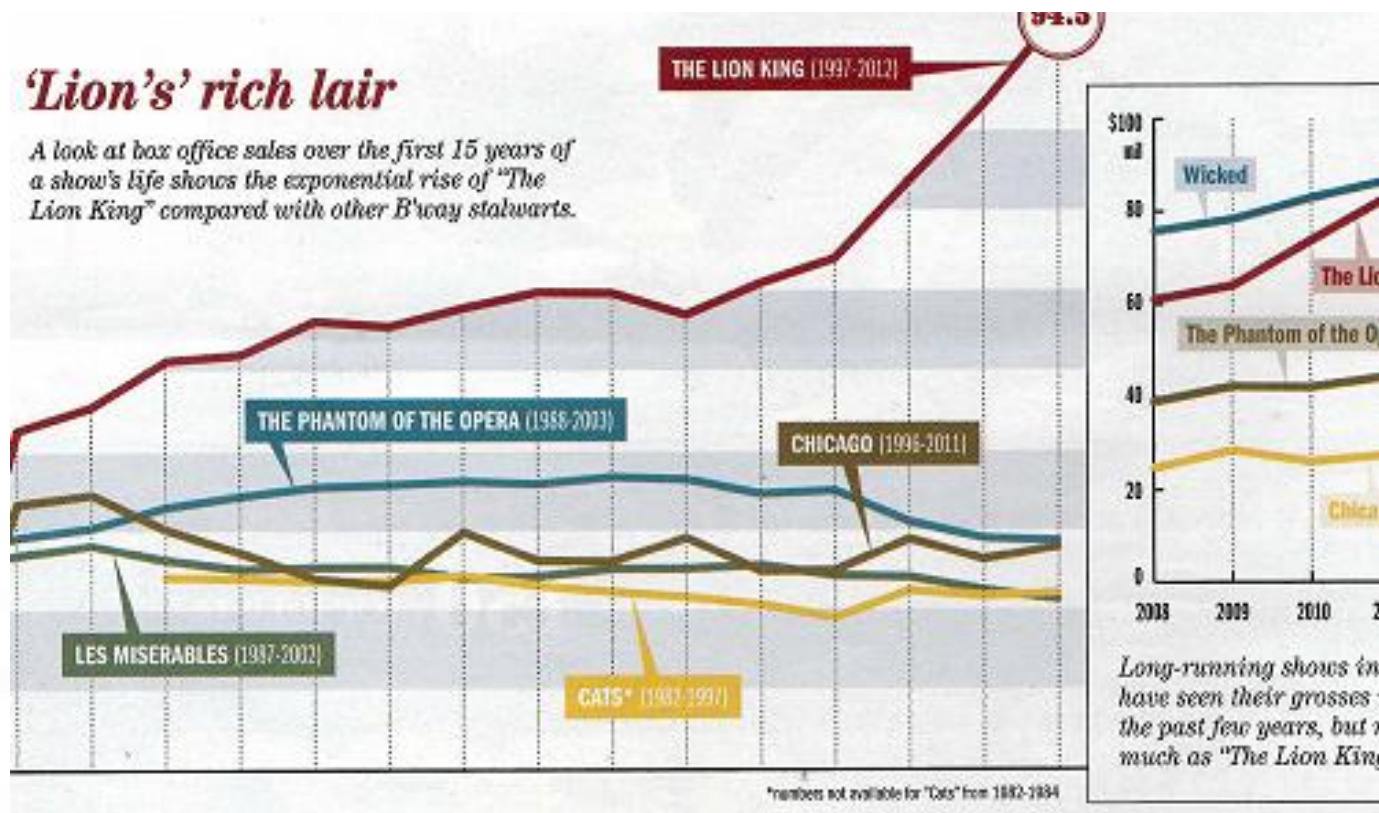
# Differential Pricing Analytics

## Why “The Lion King” Roars So Loud: Business Analytics at Disney

Wednesday, December 11, 2013 - 15:21

Robert Phillips

[Disney](#) [Revenue Management](#) [The Lion King](#)





wiseGEEK

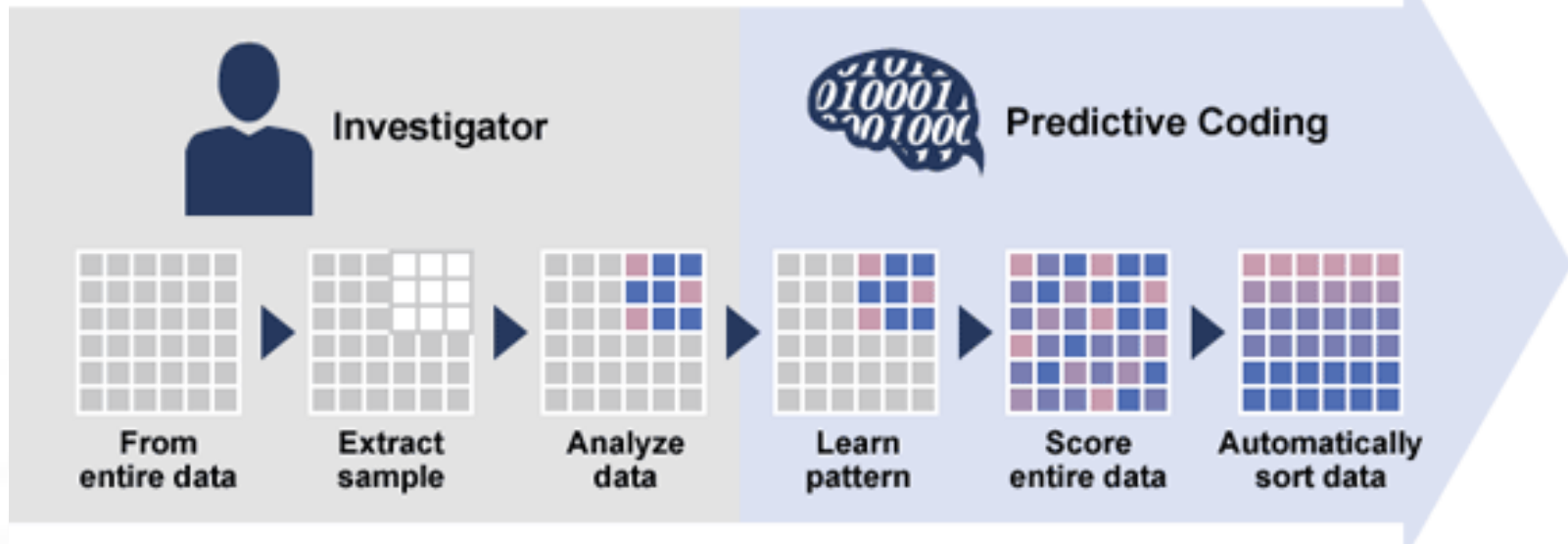
# Analytics in the Public Sphere



- Smart power generation and distribution
- Traffic management
- Waste and water treatment
- Development planning

# Analytics in the Legal Sphere

## Function Overview of Predictive Coding



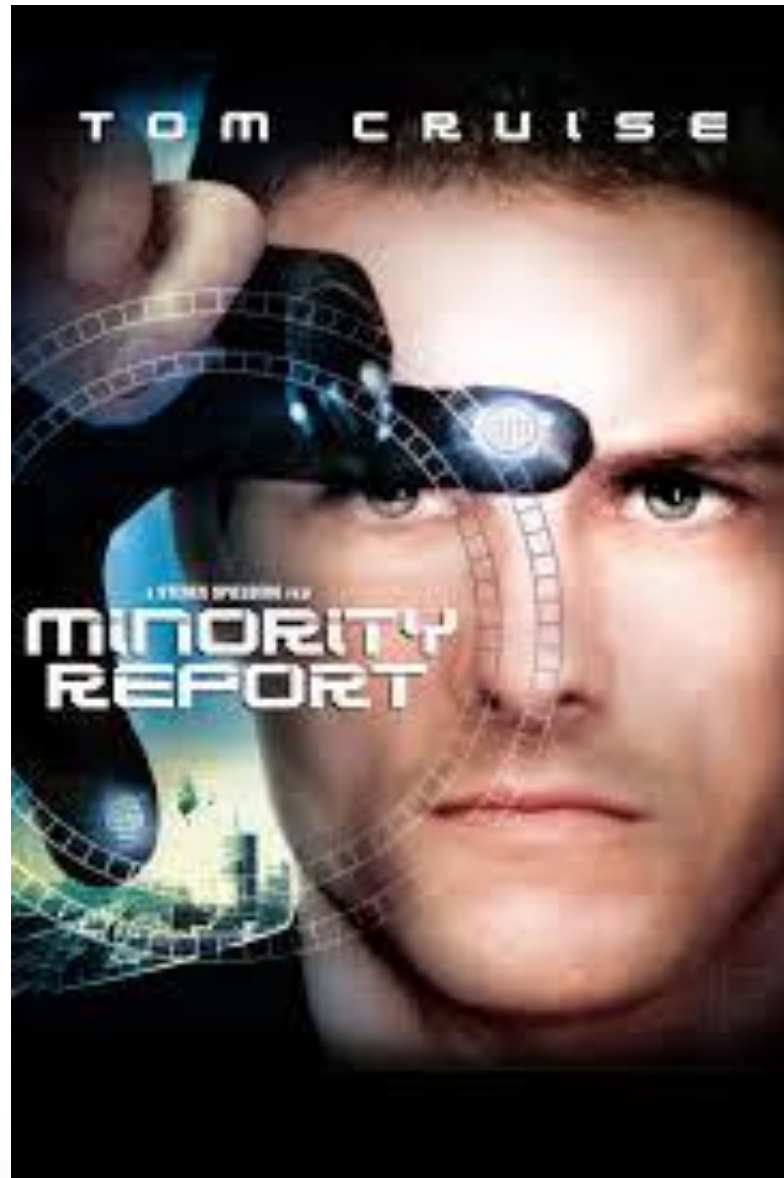
Predictive Coding is a game changer. If you want to use it, call me.

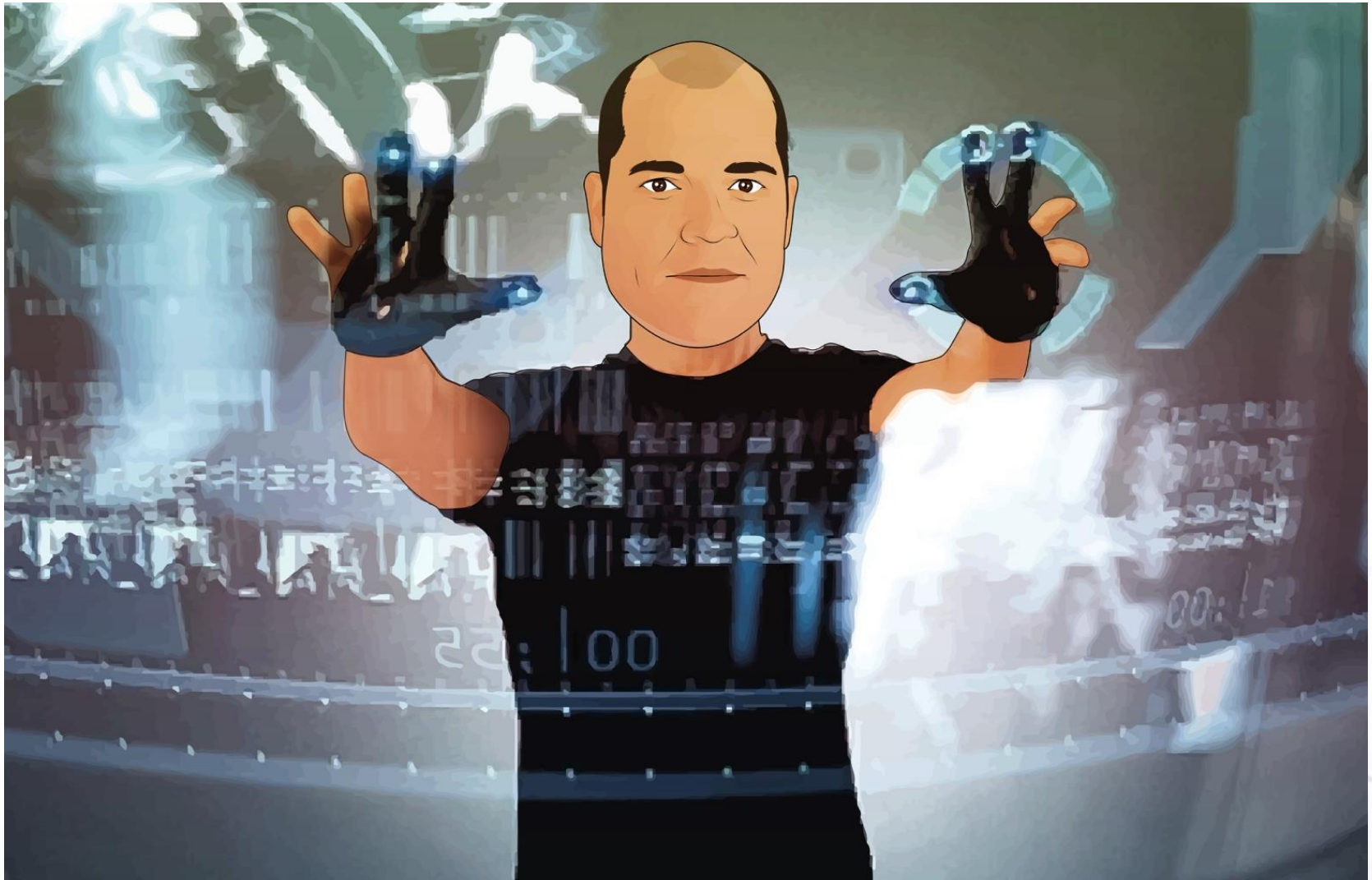
# Case Studies

- *Qui Tam* Litigation
- Drug Safety investigation
- M&A: Indemnity



# Predictive Compliance





# Our Objectives

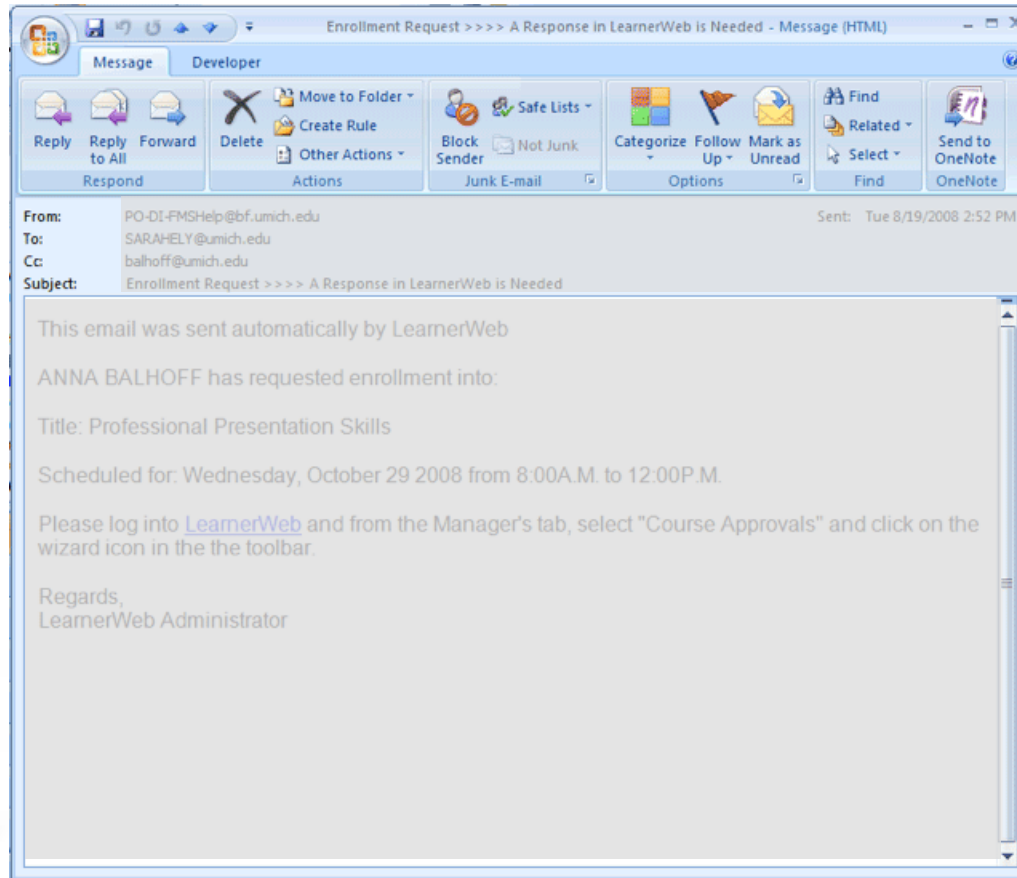
- **Identify patterns** associated with misconduct
- **Detect these patterns** in email and unstructured data
- **Develop models** to be used in an early warning system to detect and prevent misconduct *as it develops*



# The Data

Dataset Code Name	Number of Documents	Case Subject
Alpha	38,173	Financial fraud by target in acquisition
Mike	162,136	Theft of confidential information
Oscar	93,245	Fraudulent inducement of investors
Papa	964,458	Conspiracy to acquire stolen goods
Papa Bravo	9,715	Insurance fraud
Zulu	109,707	Misrepresentation of financial results

# Email Transformation



## Metadata

To  
From  
Date-Time  
Domains  
Attachments  
...

## Text

# Metadata Features

- Used many of the metadata features related to email
- Developed aggregate features
  - Internal vs. external communications vs. personal
  - Number of correspondents
  - Time of day (binned into before, during and after work hours)
  - Days of week
  - Time of year (e.g., reporting periods)
  - Binary attachment (yes/no)
  - Number of attachments
  - Size of attachments
  - Kinds of attachments
  - And many more

# Text Mining

- Distill content into words and phrases - DTM
- TF-IDF
- Sentiment analysis
- Sentiment polarities
- Custom domain dictionaries

***Result: 1,300+ features for every email***

# DTM and TF/IDF

double legal falsify stress deal fan loose Bribe WTF  
ransack number swindle timing straits crime pinch graft later alone  
blackmail tamper tell cabal perjury fuck forward Money QT mirrors mobile  
unauthorized pot doublecross sweeten disloyal hoodwink black  
harsh Shit Make Hush circumvent hush-up duplicity Pyramid Payola  
ripoff make purloin Cover stay irregular reservation  
improper Come fishy wrong plot fib Call deception hot  
part payment Fix half storm misrepresentation god cell  
trail talking two talk clear surreptitious  
sidestep bush conspiracy shady leave predicament walk countermine  
wheel envelope around low poach laundering comfortable  
defamation trick slippery whitewash Malicious Grease Kickback prohibit  
proscrib bounds devious Friendly Special induce flimflam smoke  
cheat artifice hits submit scandal fudge frame illicit truth Fraud faced  
hustle white crook sorry get lead fast Hack Steal  
loot

# Top Targeted Terms

## Oscar

Term	Documents with hits
"Call me"	43,048
"Come by my office"	23,510
"not legal"	22,228
(Take w/2 (him OR her OR them) w/2 down)	17,612
"Keep it between"	16,650
get around	11,597
prohibit*	11,158
"under the table"	9,789
hold up	9,766
"talk to me"	9,166
"Cover up"	8,675
((They OR she OR he) w/2 (owe* it to me))	7,055
"lead on"	6,252
"stay out"	5,368
breach	5,106
"don't tell"	4,064
Violation	3,788
unauthorized	3,631
"This is wrong"	3,072

## Alpha

Term	Documents with hits
Trouble	64,433
"Call me"	63,511
wrong	43,201
breach	35,379
dodge	30,026
((("Do not" OR "Don't") w/2 send)	28,937
frame	28,281
Violation	27,904
((("do not" or "don't" w/2 tell)	26,289
unlawful	19,532
"not legal"	19,288
Fraud	16,819
Secret	15,669
Forge	14,163
criminal	13,478
stress	11,622
Misconduct	10,878
Inappropriate	10,775
forbidden	10,250

# Sentiment Analysis

Motivation

Change

Power

Rectitude

Respect

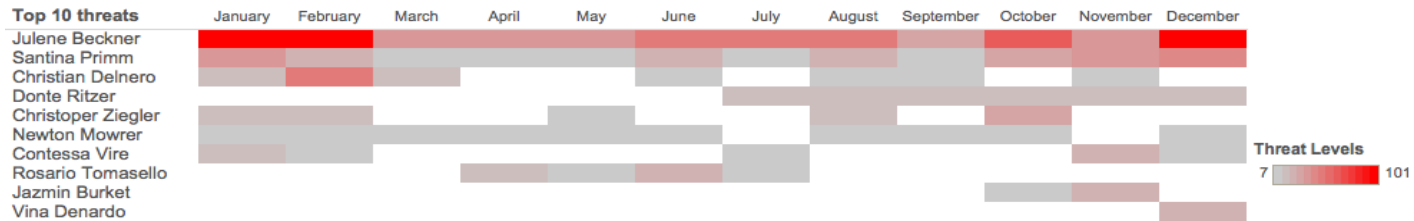
Conflict

Evasion

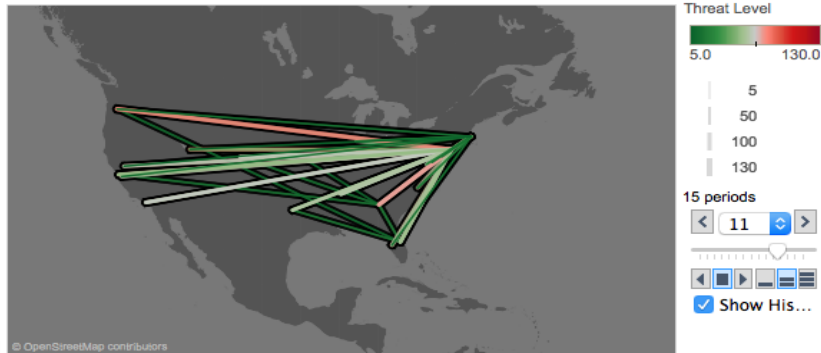


# Visualizing Misconduct

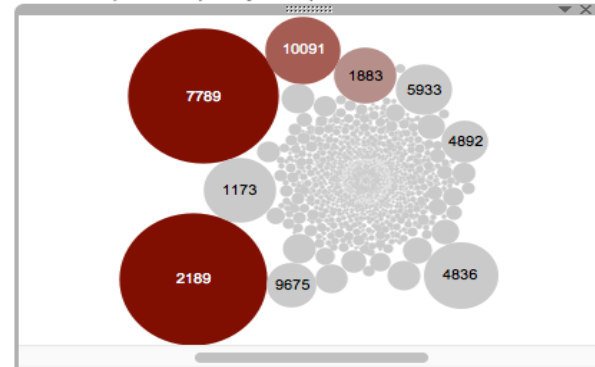
## Apocalyitics: Executive Dashboard



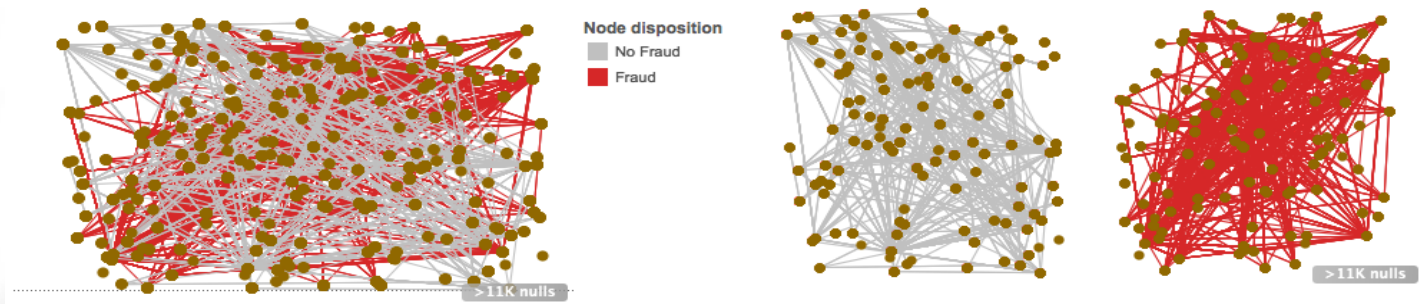
### Top threat evolution - time and location



### Current top threats (anonymized)



### Cummulative network graph of all email communication over review period (side by side view - fraud nodes not separated vs separated)





**Lawyers Who Can't Leverage Information Are **Extinct**:  
They Just Don't Know it Yet**



# Questions?

**Bennett B. Borden**

*Chief Data Scientist and Chair, Information  
Governance and eDiscovery*

**Drinker Biddle & Reath LLP**

**1500 K Street, N.W.**

**Washington, DC 20005-1209**

**(202) 230-5194 office**

**[Bennett.Borden@dbr.com](mailto:Bennett.Borden@dbr.com)**

**[www.drinkerbiddle.com](http://www.drinkerbiddle.com)**