

Predictive Coding: The Future of eDiscovery

Morgan Lewis
technology
may-rathon



presenters

Stephanie A. “Tess” Blair

Scott A. Milner

May 15th, 2012



Introduction

Please note that any advice contained in this presentation is not intended or written to be used, and should not be used, as legal advice.

Overview

- The eDiscovery Problem
- Evolution of a Solution
- Predictive Coding
- Defensibility
- Getting Started
- Early Results

The eDiscovery Problem

The eDiscovery Problem



- Volume
 - The Digital Universe doubles every 18 months
 - Corporate data volumes increasing
 - 98% of all information generated today is stored electronically
 - **2010: 988 Exabytes**
(1 Exabyte = 1 trillion books)

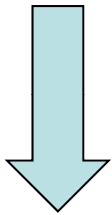
The eDiscovery Problem



- Expense
 - eDiscovery market expected to hit \$1.5 billion by 2013
 - eDiscovery can consume 75% or more of litigation budget
 - Primary cost driver is volume of information subject to discovery

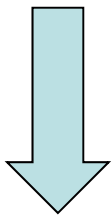
Evolution of a Solution

- Early focus on driving down cost of labor



- *Traditional Associates \$\$\$*
- *Contract Attorneys \$\$*
- *LPO \$*

- Current focus on driving down volume of data subject to discovery



- *Key words*
- *Analytics*
- *Predictive Coding*



Morgan Lewis

Evolution of a Solution

Linear Review	Limited NonLinear Review	Relevance/Priority-Centric Review
Traditional Model <ul style="list-style-type: none"> • Custodian driven 	2nd-Generation Model <ul style="list-style-type: none"> • Keyword/topic driven 	3rd-Generation Model <ul style="list-style-type: none"> • Substance driven; computer expedited
Expensive <ul style="list-style-type: none"> • False positives • Lack of context • Manual - slow • Keyword driven • No prioritization • Multipass required 	Less Expensive <ul style="list-style-type: none"> • Docs/hr improved • Limited context • Mostly manual - faster • Keyword focused • No prioritization • Multipass still required 	Least Expensive <ul style="list-style-type: none"> • Predictive Analytics™ • Domain & relevance • Technology assisted - fastest • Meaning based • Docs prioritized • Multipass optional
Unnecessary Risk <ul style="list-style-type: none"> • Many false negatives • Many false positives • No consistency • Contract attorneys 	Unnecessary Risk <ul style="list-style-type: none"> • Many false negatives • Many false positives • Limited consistency • No learning 	Limits Risk <ul style="list-style-type: none"> • Identifies false negatives • Identifies false positives • Maximum consistency • Expert driven

Predictive Coding Defined

Predictive Coding Defined

- What it is **NOT**:
 - Artificial intelligence
 - The end of attorneys reviewing documents
 - Perfect, but it is far superior to human-only, linear review

Predictive Coding Defined

- It is also **NOT**:
 - Keyword or search-term filtering
 - Near duplicates, email threading
 - “Clustering”
 - Concept groups
 - Relevancy ratings

Predictive Coding Defined

- So, what is it?
 - Computer-Assisted Review
 - Iterative, Smart, Prioritized Review
 - Faster
 - More Accurate
 - Less Expensive

Predictive Coding Defined

- Other Benefits
 - ECA
 - Quality Control
 - Privilege Analysis
 - Inbound Productions

Predictive Coding Workflow

Step 1

Predictive Analytics™ to
Create Review Sets

Human Review

Step 2

System Training on
Relevant Documents

Computer Suggested

Step 3

Human Review of
Computer Suggested

Adaptive ID Cycles
(Train, Suggest, Review)

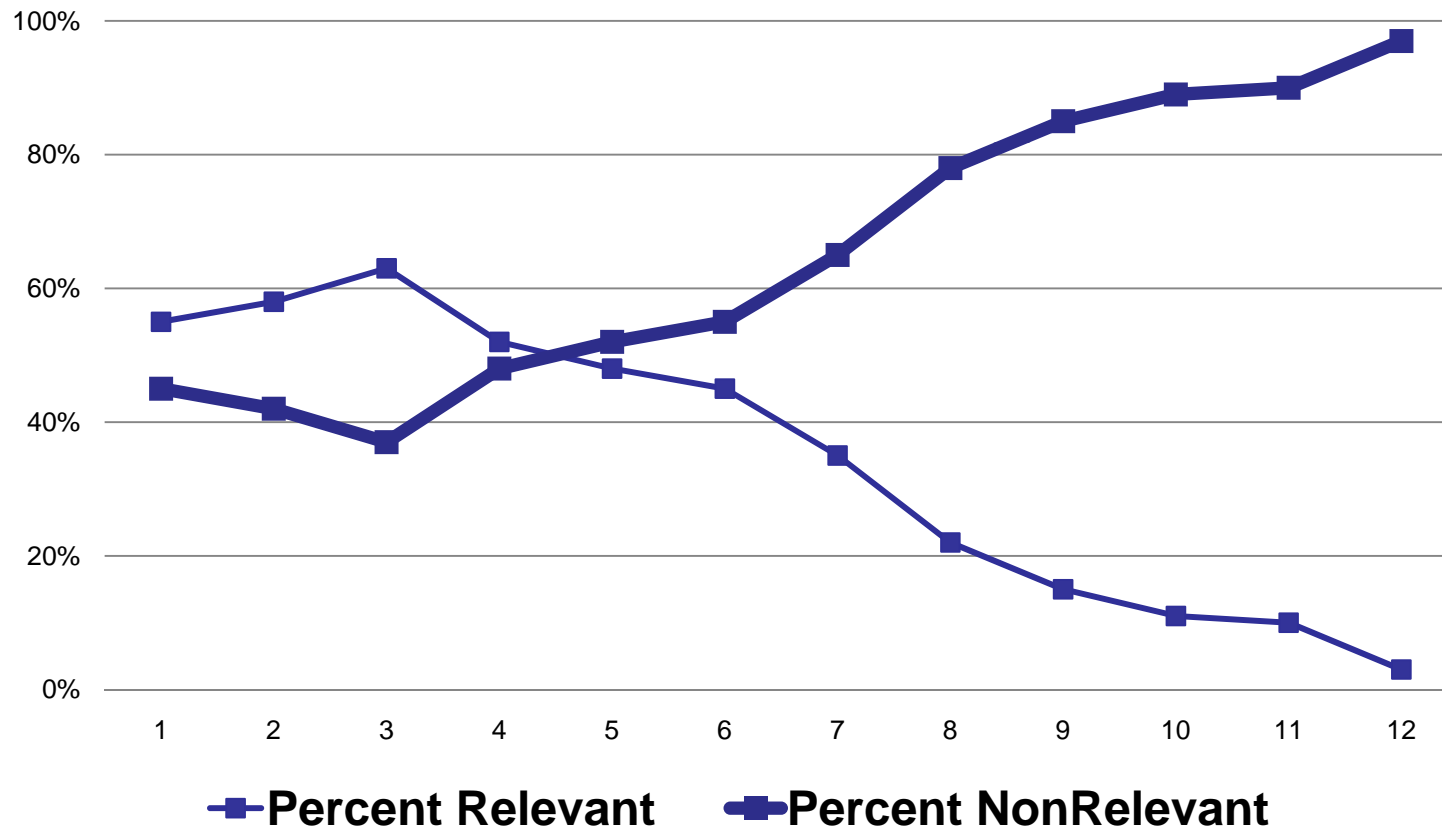
Step 4

Statistical Quality-
Control Validation



Iteration Tracking: When Are We Done?

Training Iteration Analysis

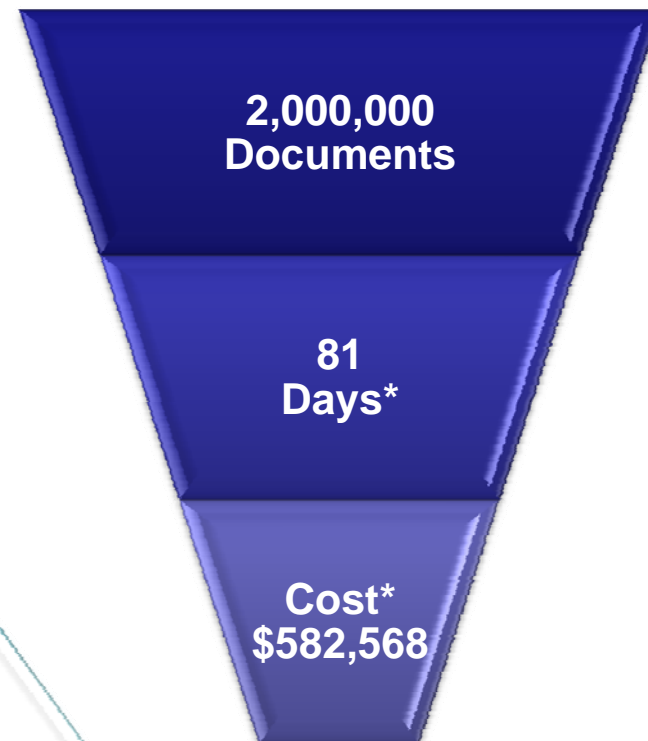


Hypothetical: Human Review vs. Predictive Coding

Linear Review



Predictive Coding



**Predictive
Coding Savings
\$1,053,796**

*Required only 35% of the collection to be reviewed.

Defensibility

Defensibility

- Defensibility
 - Predictive coding not at issue – Humans review and determine relevancy of computer-suggested documents assisted by Predictive Coding – No “black box”
 - For documents not reviewed – Issue is sampling
 - Statistical sampling widely accepted – scientific method supported by expert testimony
- Disclosure
 - Split emerging within profession on disclosure
 - Whether and when to disclose use of Predictive Coding
 - What to disclose

Defensibility

- Defensibility (cont.)
 - Case law growing on the use of sampling techniques
 - Zubulake v. UBS Warburg, LLC, 217 F.R.D. 309 (S.D.N.Y. 2003)
 - *Court accepted the use of sampling due to the prospect of having to restore thousands of archived data tapes.*
 - Mt. Hawley Ins. Co. v. Felman Prod. Inc. 2010 WL 1990555 (S.D. W.Va. May 18, 2010)
 - *“Sampling is a critical quality control process that should be conducted throughout the review.”*
 - In re Seroquel Prods. Liab. Litig., 244 F.R.D. 650 (M.D. Fla. 2007)
 - *Court instructed “common sense dictates that sampling and other quality assurance techniques must be employed to meet requirements of completeness.”*

Defensibility

- Defensibility (cont.)
 - Endorsement by legal community (Legal Tech 2012, NYC)
 - Judge Andrew Peck and judicial endorsement
 - *October 2011 LTN Article*
 - *Order in Da Silva Moore v. Publicas Groupe et al. (S.D.N.Y 2011)*

Getting Started

Key Ingredients

- Predictive Coding requires:
 - People
 - Process
 - Technology

People

- People:
 - Experienced litigators to create and QC seed set
 - Experienced discovery attorneys to drive the predictive coding workflow, gather metrics, and measure results
 - Technicians to run the technology and manage the data

Process

- Process
 - Documented workflow
 - Process capable of being repeated
 - Quality control by attorneys
 - Process for gathering appropriate metrics
 - Level of confidence supported by statistics

Technology

- Technology
 - Few software vendors offer true “predictive coding” capability
 - Many are claiming they have this technology, but are just repackaging existing technologies with new buzzwords
 - Buyer beware

Early Results

How Morgan Lewis Uses Predictive Coding

- Increase Quality
 - Error rate reduction
 - Confidence intervals
- Enhance Service Delivery
 - Cost certainty
 - Time certainty
- Demonstrate Real Value
 - Early Case Assessment
 - Discovery cost equal to value received
- Competitive Advantage
 - Dedicated technical and legal team with expertise in predictive coding
 - Pricing competitive with all other market segments, including offshore

Case Studies

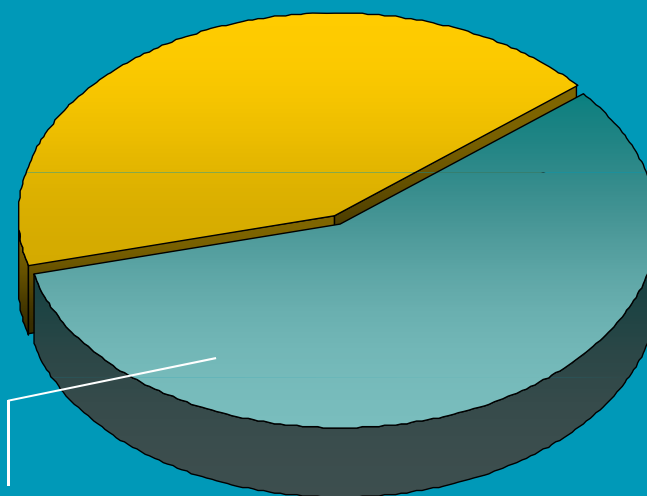
Reduction in Volume

Review and Production of ESI

Case Study 1

- Coded by computer = 57% (317,000 docs)
- Confidence interval = 95%
- Defect rate = .79% or less

552,871 total documents



57% coded by computer

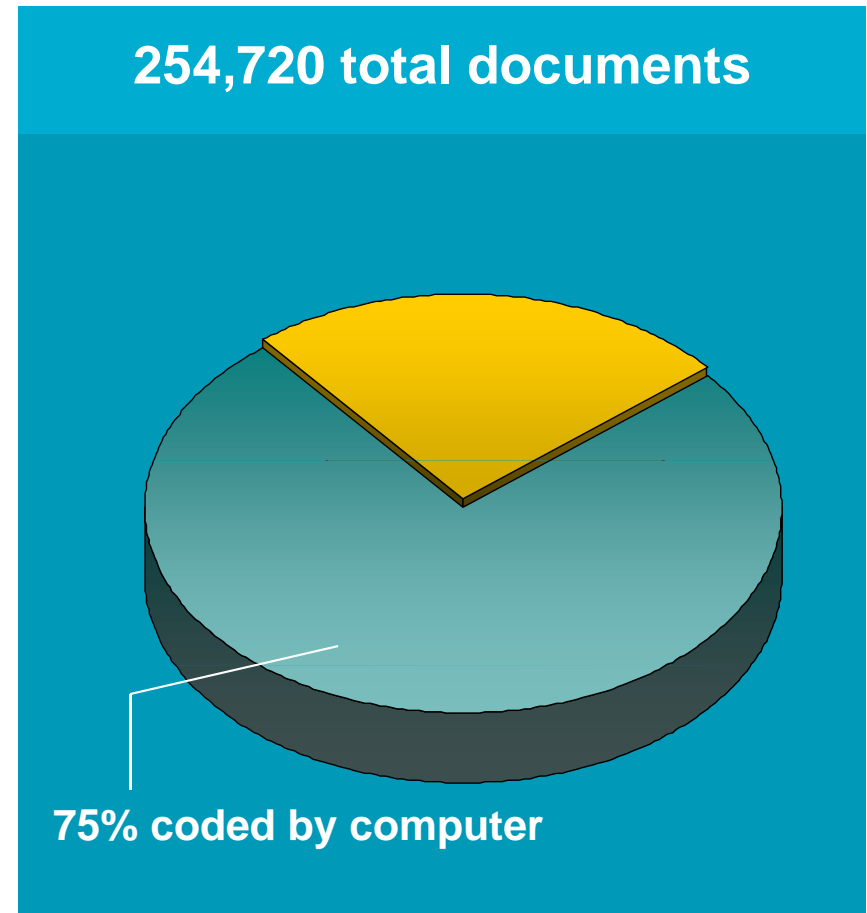
Case Studies

Reduction in Volume (cont.)

Review and Production of ESI

Case Study 2

- Coded by computer = 75% (192,000 docs)
- Confidence Interval = 95%
- Defect rate = 5% or less



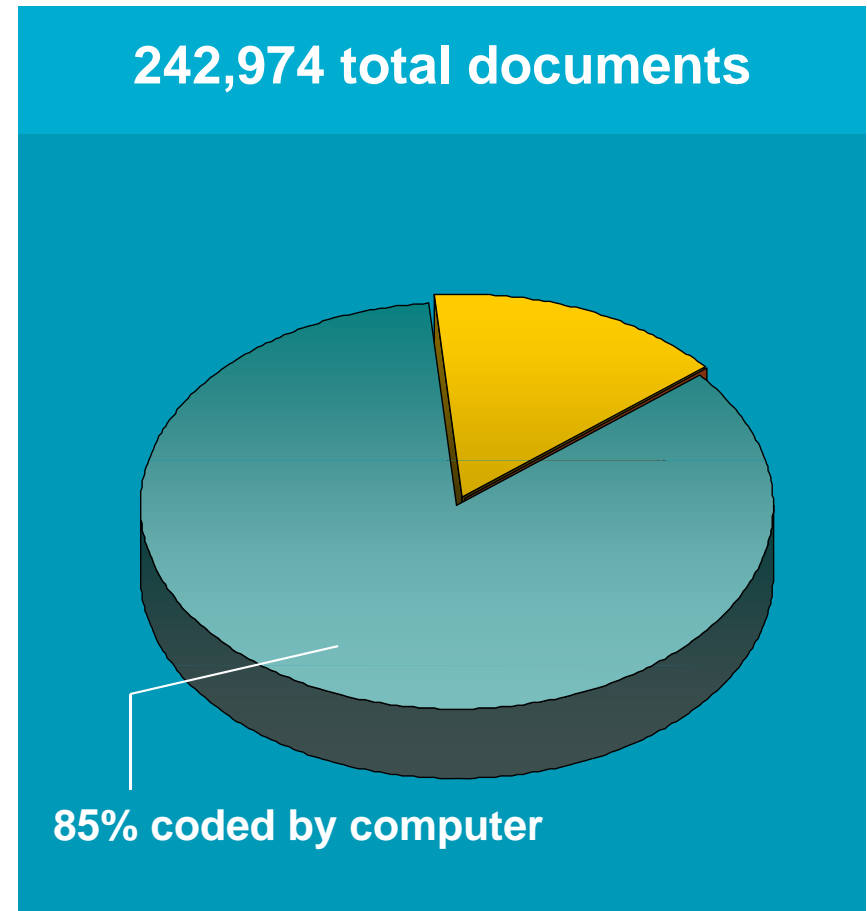
Case Studies

Reduction in Volume (cont.)

Review and Production of ESI

Case Study 3

- Coded by computer = 85% (206,000 docs)
- Confidence Interval= 95%
- Defect rate = 5% or less



Contacts

Tess Blair

Partner, Morgan, Lewis & Bockius LLP
eData Practice Group
215.963.5161
sblair@morganlewis.com

Scott Milner

Partner, Morgan, Lewis & Bockius LLP
eData Practice Group
215.963.5016
smilner@morganlewis.com

Participants



Stephanie A. Blair

Partner

Morgan Lewis

P: 215.963.5161

E: sblair@morganlewis.com



Scott A. Milner

Partner

Morgan Lewis

P: 215.963.5016

E: smilner@morganlewis.com



international presence

Beijing Boston Brussels Chicago Dallas Frankfurt Harrisburg Houston Irvine
London Los Angeles Miami New York Palo Alto Paris Philadelphia Pittsburgh
Princeton San Francisco Tokyo Washington Wilmington