2018 ICN MERGER WORKSHOP
BOS 4: TECHNOLOGY AND DATA REVIEW

Ruediger Schuett, International Counsel, U.S. Department of Justice, Antitrust Division

November 7, 2018

DISCLAIMER

The views expressed in this presentation are solely those of the presenter and do not necessarily reflect those of the U.S. Department of Justice or any component thereof.

Overview

- Role of Documentary Evidence in Merger Investigations

- Legal Framework

- Challenges in Obtaining & Reviewing Documents

- Ways to Address Challenges

- Background: Predictive Coding

Legal Framework

- Hart-Scott-Rodino (HSR) Act & HSR Rules
  - HSR Form, Items 4(c) and 4(d)
  - Voluntary Requests
  - Requests for Additional Information and Documentary Material (Second Requests)
  - Typically, both acquiror and target required to produce material

- Antitrust Civil Process Act
  - Civil Investigative Demands (CIDs) for Documentary Material
  - Issued where persons may have documents "relevant to a civil antitrust investigation" (*e.g.*, third parties in merger investigations)

Importance of Documentary Evidence

- Transaction-related documents
  - Examples: board presentations discussing deal; synergy analyses
  - → Reflect parties' views of transaction rationale and competitive effects

- Ordinary course of business documents
  - Business records (*e.g.*, win/loss reports, customer lists, pricing or sales records, bid documents, contracts)
  - → Reflect competitive position of parties and extent of competition
  - Analytical documents (*e.g.*, market studies, forecasts, surveys)
  - Planning documents (*e.g.*, business plans, budgets, strategies)
  - Documents related to prior transactions
  - → Reflect parties' views on markets, competition, future expectations

Challenges

- Ensure production of responsive documents

- Minimize production of non-responsive documents

- Ensure identification of relevant/important documents during review by Division staff

→ Huge and ever increasing amounts of documents produced in merger investigations amplify challenges

Recent Document Production Volumes

Year	Records	Images
2015	~34 million	~130 million
2016	~53 million	~134 million
2017	~87 million	~224 million

How to Cope: Focus Requests

- Tailor model Second Request to specific case
  - ➢ Several specifications in model are optional

- Limit time frame for document requests
  - ➢ Generally 2 years under model

- Limit scope of specific requests
  - ➢ Example: "documents sufficient to show" instead of "all documents"

- Limit number of custodians
  - ➢ New policy: Generally no more than 20 custodians per party

How to Cope: Cooperation with Parties

- Frequent communications regarding document collection, review, and production
  - ➢ Detailed specifications in Second Requests/CIDs
  - ➢ Questionnaire on Electronically Stored Information
  - ➢ Calls to discuss potential issues and modifications

- Written confirmation/agreements on major steps
  - ➢ Written confirmation of Second Request/CID modifications
  - ➢ Timing agreement (*e.g.*, custodians, priorities, schedule)
  - ➢ Agreement on review process (search terms, predictive coding)
  - ➢ Deduplication agreement

How to Cope: Improve Review Process

- Given size of productions, page-by-page review of full productions not realistic
  - Experiment with using temporary contract attorneys to increase review volume, but difficult to fit into Division structure

- Instead, focused review to identify important documents related to critical aspects of case
  - Requires careful planning: What are critical issues? Who are the important custodians? What is the relevant date range?

- Use of new review tools and techniques
  - Filtering
  - Search/Analytical Tools

Lessons Learned

- Specialized staff needed to handle technical aspects

- Be realistic about resources/time for investigation

- Coordination with parties crucial

- Utilize review platforms to deal with large productions

- Approach outsourcing with great caution

Questions?

Backup:

Technology-assisted Review (Predictive Coding)

1. What is predictive coding?

2. How does it compare to traditional document review?

3. How to negotiate a predictive coding protocol?

Predictive Coding

1. What is predictive coding?

Predictive Coding is:

The use of **machine learning** to:

- categorize a collection of documents

- based on lawyer's review of a **subset** of the collection

- via an **iterative** process

- **validated** through defensible statistical **sampling** techniques

Predictive Coding is Really Like:

- A binary choice:
  - Responsive
  - Not Responsive

- Categorization made by subject matter expert

Predictive Coding is a **Process**:
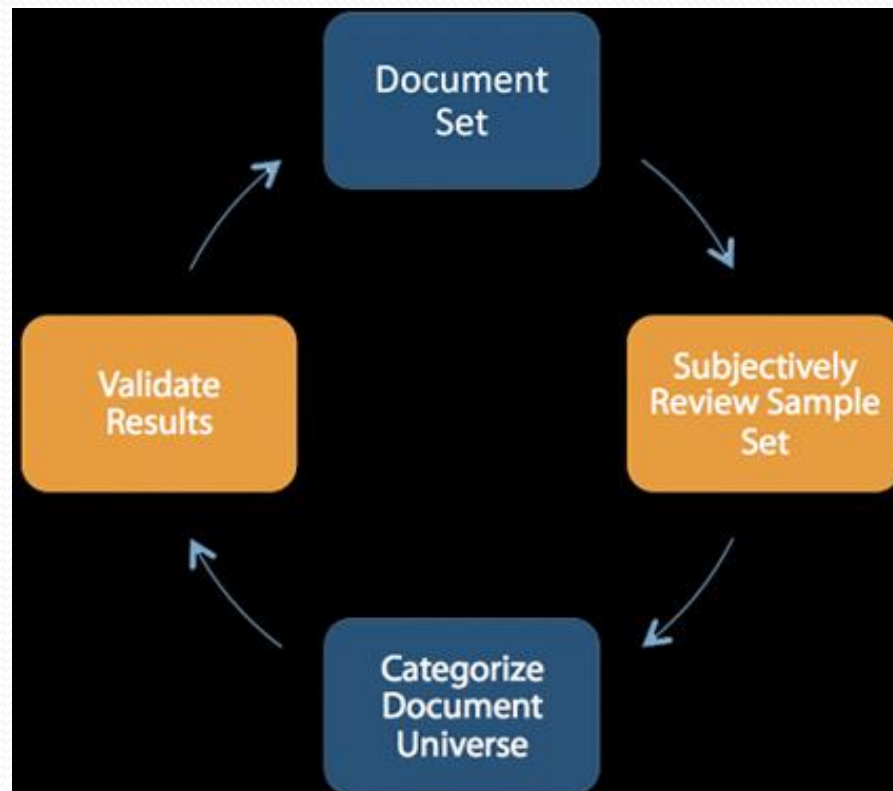
Five Key Steps:

- Subject matter experts code documents in a representative sample (training set) as responsive or non-responsive

- Software "learns from" categorized documents to apply the same categorization to all documents in the collection

- Software refines its categorization based on additional samples categorized by the subject matter experts

- The accuracy of the categorization is measured statistically to assess its effectiveness

- After entire collection is categorized, experts QC additional samples to ensure the validity of the overall classification

Predictive Coding

1. What is predictive coding?

2. **How does it compare to traditional document review?**

Traditional Document Review

- Manual (Linear) review of documents by a team

- Batches of documents assigned for review based on combination of:

  - Keyword searches

  - Custodian

  - Dates

- Even the best organized manual (linear) document review finds less than 50% of responsive documents

  ➤ Often, far less

- Why are the results so poor?

  ➤ Tired

  ➤ Inconsistent

  ➤ No effective means to QC initial review

Advantages of Predictive Coding

- Allows to review large document collections more quickly with a smaller team

  - ➢ Document collections are getting larger

  - ➢ Available resources are not

- More relevant; fewer irrelevant documents

- Because it's more accurate, potentially fewer disputes

Predictive Coding

1. What is predictive coding?

2. How does it compare to traditional document review?

3. **How to negotiate a predictive coding protocol?**

Negotiation should Address:

- Is this an appropriate case for predictive coding?

- Collection & Workflow

- Privilege

- Validation

- Written Agreement + Communication

Is this an appropriate case?

- How many documents will be produced?
  - ~100,000 documents minimum

- For especially large cases, Division may propose using predictive coding

- Trying to "insert" predictive coding into an existing discovery protocol can lead to errors, disagreements

Where predictive coding fails

- Numbers

- Graphics/photos

- Foreign languages

- Audio Files

- OCR

Predictive coding also fails

Phone Calls Between Publisher Defendant CEOs in December and January

HarperCollins — Brian Murray

SIMON & SCHUSTER — Carolyn Reidy

MACMILLAN — John Sargent

Penguin — David Shanks

hachette BOOK GROUP — David Young

29 March 2007 Trader A asked what Japanese yen Libor submission UBS was going to set. Manager A: "Too early to say yet … prob[ably] .69 would be our unbiased contribution … as i said before - i dun mind helping on your fixings, but i'm not setting libor 7bp away from the truth i'll get ubs banned if i do that, no interest in that".

14 July 2009, Trader A requested of a broker in an electronic chat a higher Libor rate: "SUPERMAN … BE A HERO TODAY." Broker F said: "ill try mate … as always."

15 July 2009 in an electric chat Trader A requested: "3m and 1m unch [i.e. unchanged]". Trader A also inserted an extract of another electronic chat with Broker A of Broker Firm A in which Broker A said: "Putting the captain caos [sic] outfit on as we speak".

21 July 2009, Trader A contacted Broker E of Broker Firm B who advised him to make small changes to his Libor submissions: "If you drop your 6M dramatically on the 11th mate, it will look v fishy, especially if [Panel Bank 5] and [Panel Bank 2] go with you. I'd be v careful how you play it, there might be cause for a drop as you cross into a new month but a couple of weeks in might get people questioning you." Trader A replied: "Don't worry will stagger the drops …"

Collection & Workflow

- To be addressed upfront

- Document collection:
  - Are search terms being used? – Consider alternatives (custodians, date range, filtering)
  - Self-collection issues?

- Workflow:
  - Depends on platform
  - Who is doing the coding? – Subject matter expert required
  - Reporting metrics
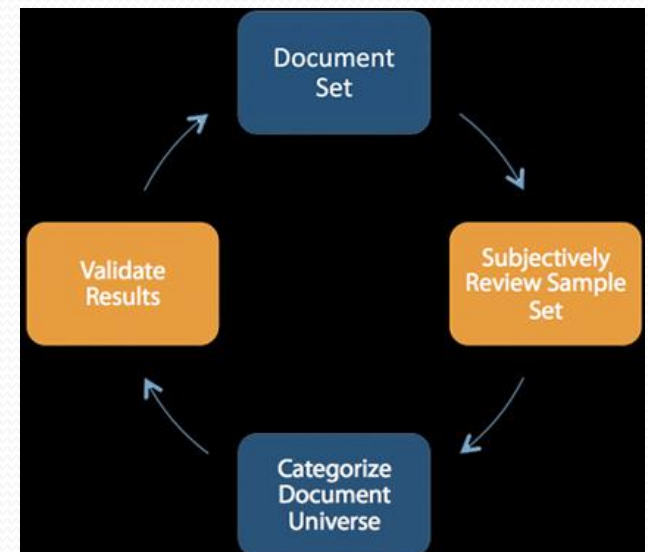
Privilege Review: Continuing Conundrum

- Parties usually demand page by page privilege review
  - ➢ Significantly limits cost/time savings of predictive coding
- How to address in the investigative context?
- Guard against "stealth" linear review

Validate the Results

- Most controversial issue

- Different approaches:

  - Sharing "training" or "seed" set

  - Joint training of software

  - Review sample of non-responsive

  - Require disclosure of metrics

  - "Trust us"

Written Agreement + Communication

- Predictive Coding Protocol agreed on <u>in advance</u>

  ➢ Drafted by Division staff

- Protocol emphasizes importance of communication between lawyers, technical staff, and vendor

  ➢ Technology is new – expect the unexpected

  ➢ Division affirmatively asks whether any strange results

  ➢ Easier to address problems before production is made