

Xerox Litigation Services

E-Discovery and Technology-Assisted Review



Amanda Jones, Senior Search Consultant

What is Discovery?

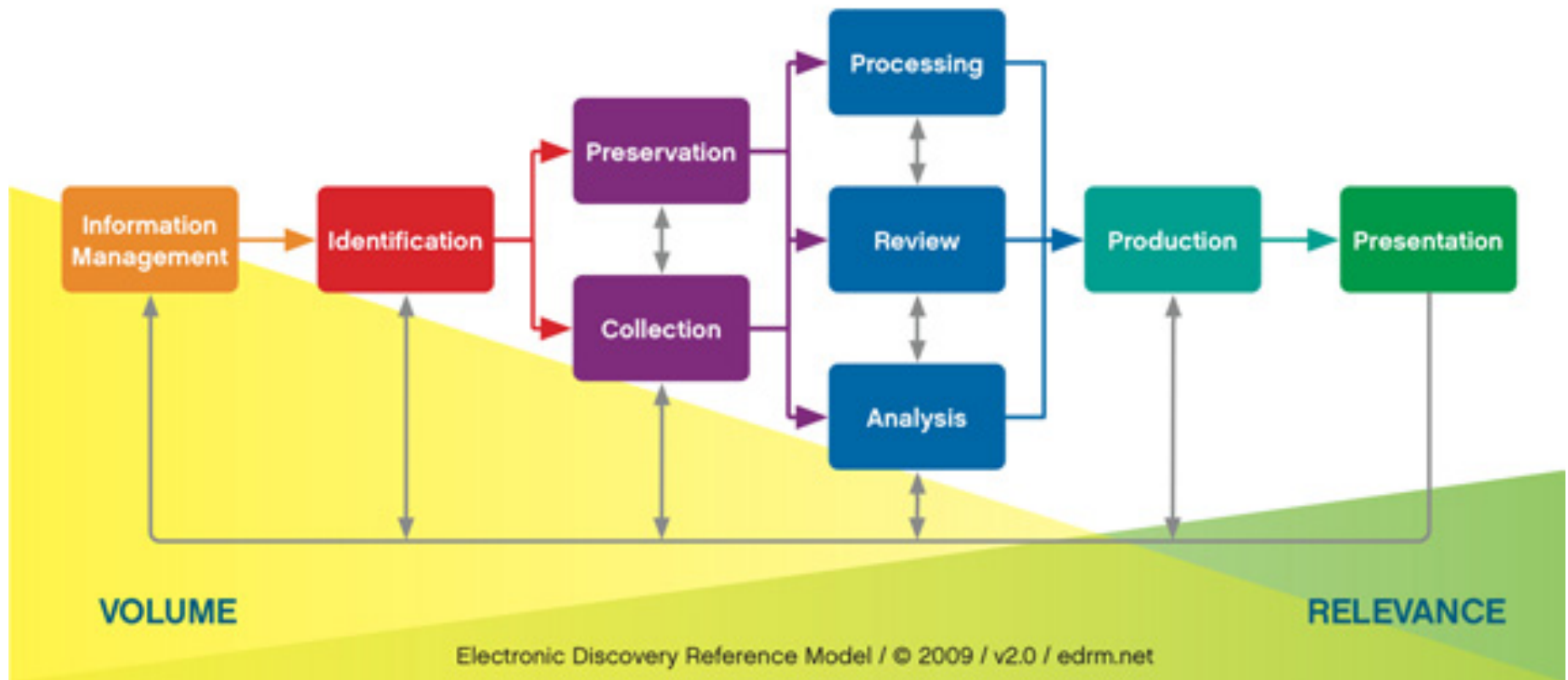
The pre-trial phase of litigation during which parties obtain evidence from one another.

- Discovery is important. Both parties learn the strengths and weaknesses of their cases.
- Most trials – e.g. the Microsoft vs. Real Audio patent infringement matter – settle before they ever reach trial because of Discovery.

What is E-Discovery?

- Electronic discovery (e-discovery) is simply discovery that involves electronically stored information (ESI).
- E-discovery has grown in importance, due to the skyrocketing volume and variety of ESI.
 - Terabytes of data
 - Millions of documents
 - Many document sources and types: emails, spreadsheets, pdfs, IMs, images, audio, etc.

Electronic Discovery Reference Model



E-Discovery – Strategic Considerations

- Document review is traditionally the most expensive aspect of discovery.
 - Manual review requires large teams of skilled laborers.
- Legal teams can exploit this fact to gain strategic advantages.
 - Maximize the opponent's review burden, while minimizing your own.

E-Discovery – The Challenge

- High Precision and High Recall Imperative
 - Legal obligation to produce *all* potentially relevant material
 - Obligation to client to produce *only* potentially relevant material
- Preserving both fair access to justice and high quality standards demands new strategies for document review.

E-Discovery – Technology Adoption in a Risk-Averse Industry

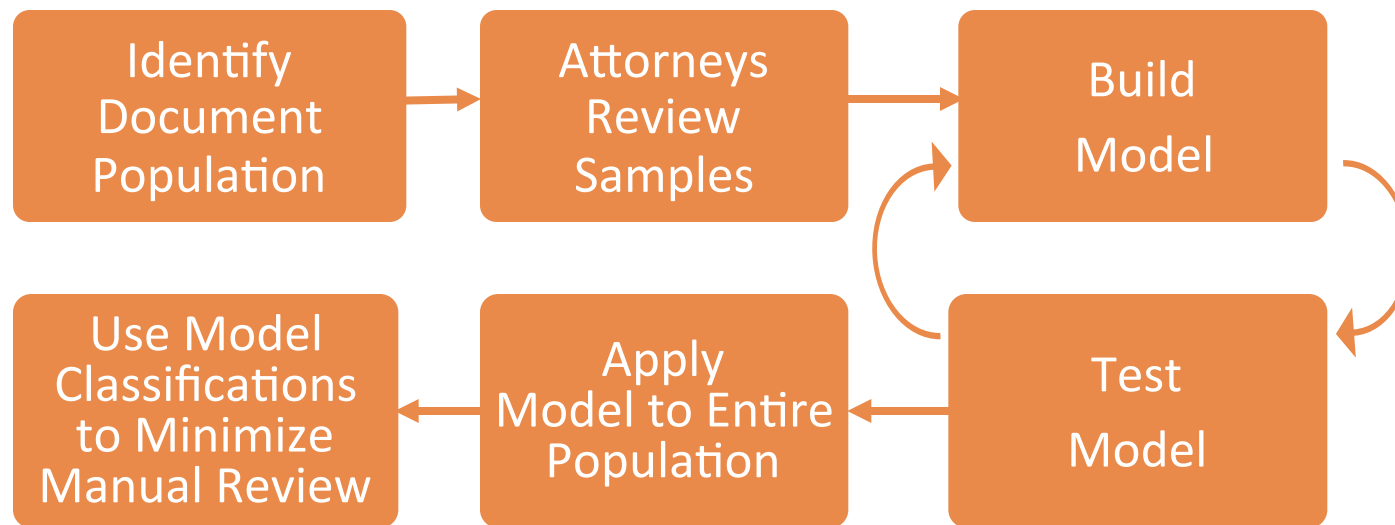
- Attorneys, however, tend to be risk averse.
- What contributes to risk aversion?
 - Attorney responsibility for certification
 - Threat of sanctions
 - Reliance on precedence
 - Fear of the “black box”
 - Defensibility

E-Discovery – Technology Adoption in a Risk-Averse Industry

➤ What mitigates risk aversion?

- Legal Concepts of “Reasonableness” and “Proportionality”
- Emerging case law and evolving rules for discovery
- Academic and industry thought leadership
- Emphasis on process rather than technology
- Bottom Line: *Need*

E-Discovery – Technology-Assisted Review Basics



E-Discovery – Technology-Assisted Review Variations

➤ Which algorithms are used?

- Naïve Bayesian Classifiers
- Latent Semantic Indexing
- Support Vector Machines
- Probabilistic Latent Semantic Analysis
- Sparse Logistic Regression
- Combinations of the above

E-Discovery – Technology-Assisted Review Variations

- How is the data processed prior to application of the algorithm(s)?
 - Generally full text is extracted from each individual document (not full families).
 - Some systems leverage metadata information, but many do not.
 - Most systems will impose some text filtering, but the specifics will vary widely.

E-Discovery – Technology-Assisted Review Variations

- How are samples chosen for training and/or testing?
 - Pure random sampling
 - Active learning or other non-random sampling
 - Combination of random and non-random

E-Discovery – Technology-Assisted Review Variations

- How is performance measured?
 - Sometimes rigorously with formal Precision and Recall metrics
 - Sometimes, not so much...

E-Discovery – Technology-Assisted Review Variations

- Is the application offered as a service or as a push-button DIY function?
 - Tremendous variability – offerings run the gamut

E-Discovery – Technology-Assisted Review Variations

➤ How are the classification results utilized?

- Binary decision-making
- Prioritization
- Quality Control
- Any combination of the above
- Alone or in combination with other review-streamlining strategies

