



IBM 2010 新锐洞察高峰论坛

心·睿 掌控制胜先机





基于内容的eDiscovery与分析

Name: Paul HC Cheung

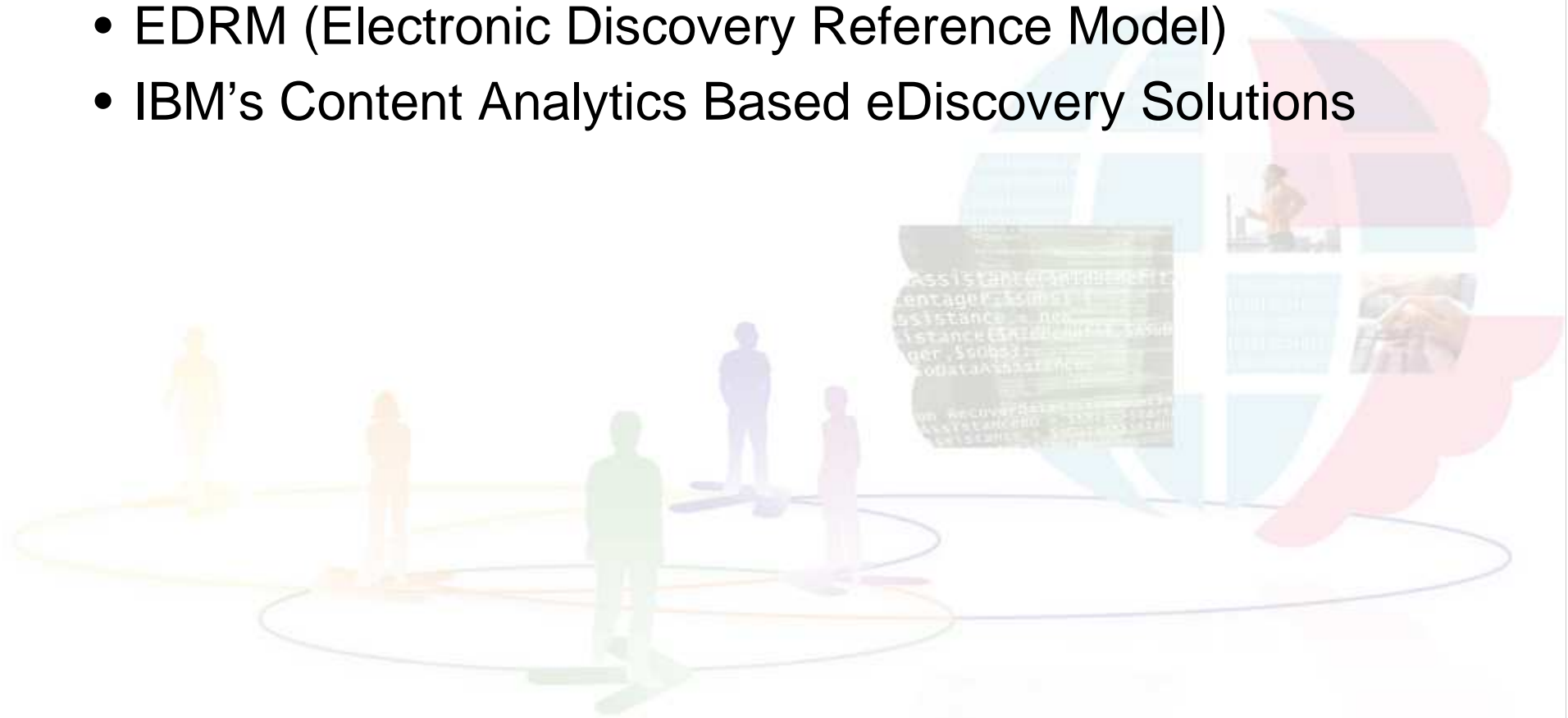
Department: Software Group

Title: Consulting IT Specialist



Agenda

- eDiscovery and FRCP (Federal Rules of Civil Procedure)
- EDRM (Electronic Discovery Reference Model)
- IBM's Content Analytics Based eDiscovery Solutions





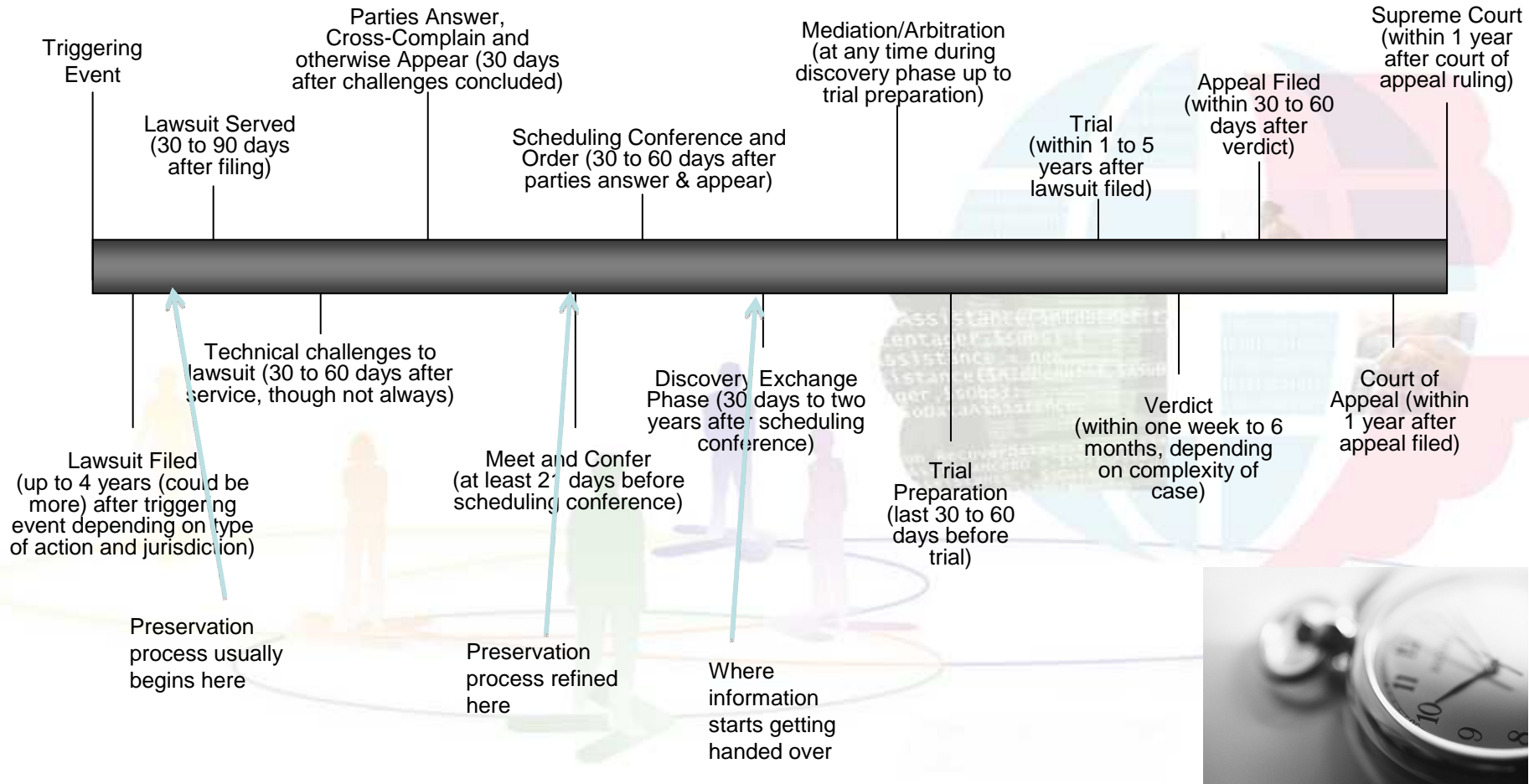
Federal Rules of Civil Procedure

- Chapter I – Scope of FRCP
- Chapter II – Commencement of Suits
- Chapter III – Pleading and Motions
- Chapter IV – Parties
- **Chapter V – Discovery**
 - Discovery is the process whereby civil litigants seek to obtain information both from other parties and from non parties (or third parties). . . .
- Chapter VI – Trial
- Chapter VII – Judgment
- Chapter VIII – Provisional and Final Remedies
- Chapter IX – Special Proceedings





Timeline of a Typical Case





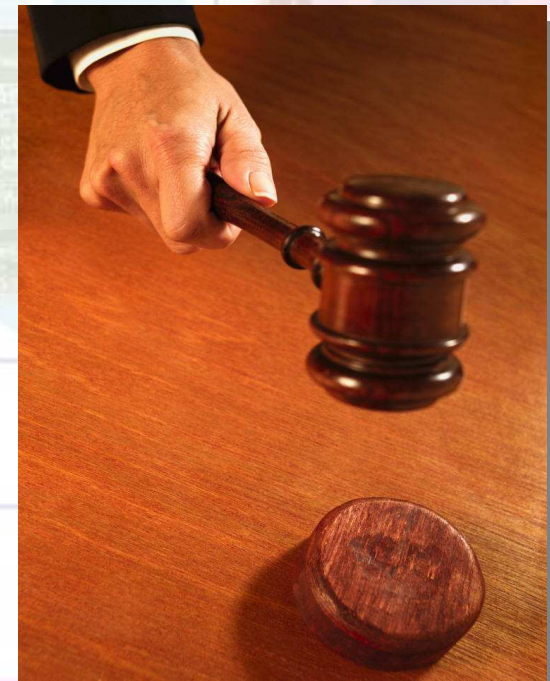
eDiscovery: The Legal Landscape

- Over the past five years, the concept of electronically stored and searchable information has led to a significant evolution in the law governing discovery in the U.S.
- New Federal Rules Amendments affecting eDiscovery took effect on December 1, 2006
- The FRCP Amendments represent an attempt to harmonize emerging local court practices and codify the law, rather than an effort to radically change the law that has developed over the past five years
- Important developments involving the standards applicable to the use of search terms (e.g., Victor Stanley v. Creative Pipe, U.S. v. O'Keefe)
- Headline-grabbing sanctions decisions involving the duty of counsel to have adequately searched for responsive documents (e.g., Qualcomm v. Broadcom)



Electronic Discovery Goes Beyond Litigation

- eDiscovery enables multiple use cases, including handling:
 - Evidence finding for utilization in the legal process
 - Regulatory matters
 - Investigations
 - Internal Investigations
 - Governmental Investigations
 - Employee disputes
 - Corporate Governance
 - Corporate Transactions
 - Operational Control





And eDiscovery is a Global Concern

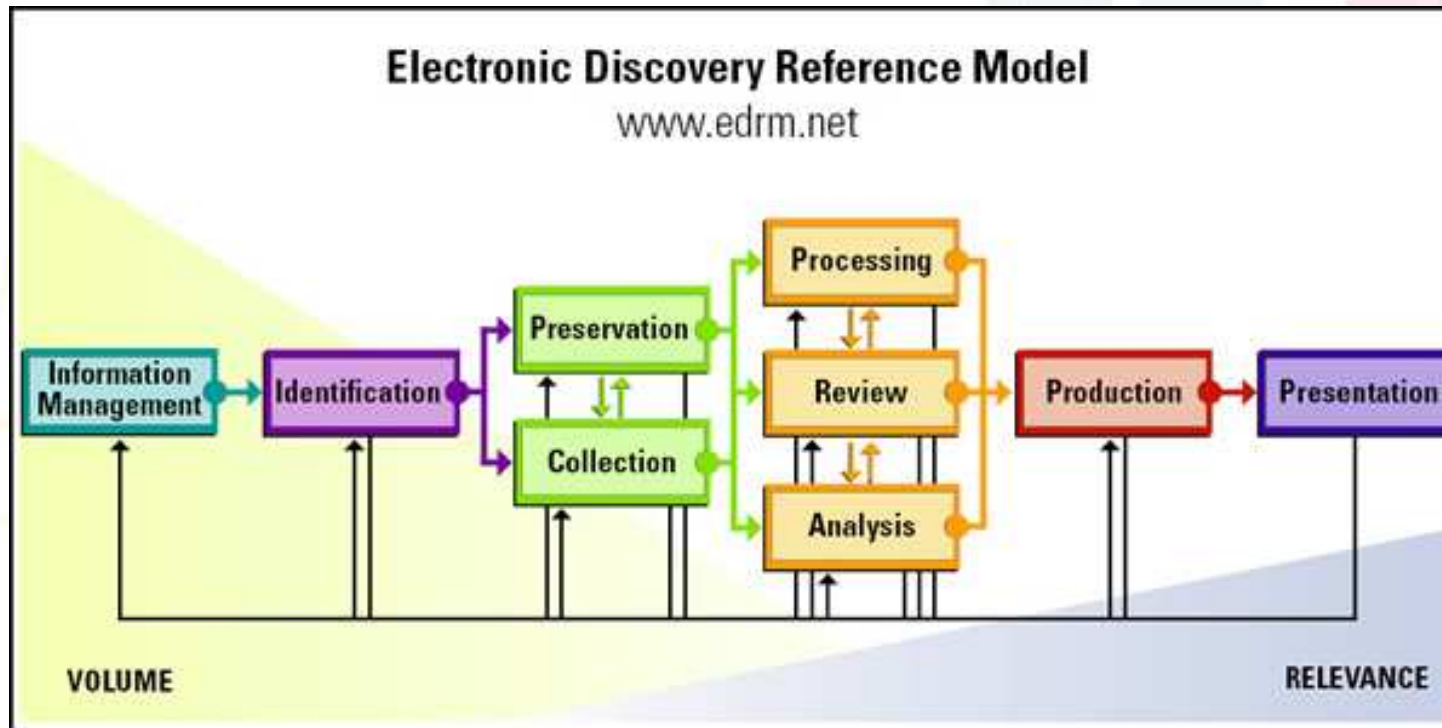
- Primary driver for eDiscovery is in common law jurisdictions like the US
- But flattened global boundaries make eDiscovery a concern for any multinational/global organization
 - Litigation resulting from business in US
 - Tactically transacting litigation in US to leverage discovery – IP & patents
 - Emerging discovery requirements in non-US litigation
 - Singapore - Practice Law number 3
 - Internal & regulatory investigations
- International eDiscovery brings additional challenges, including privacy





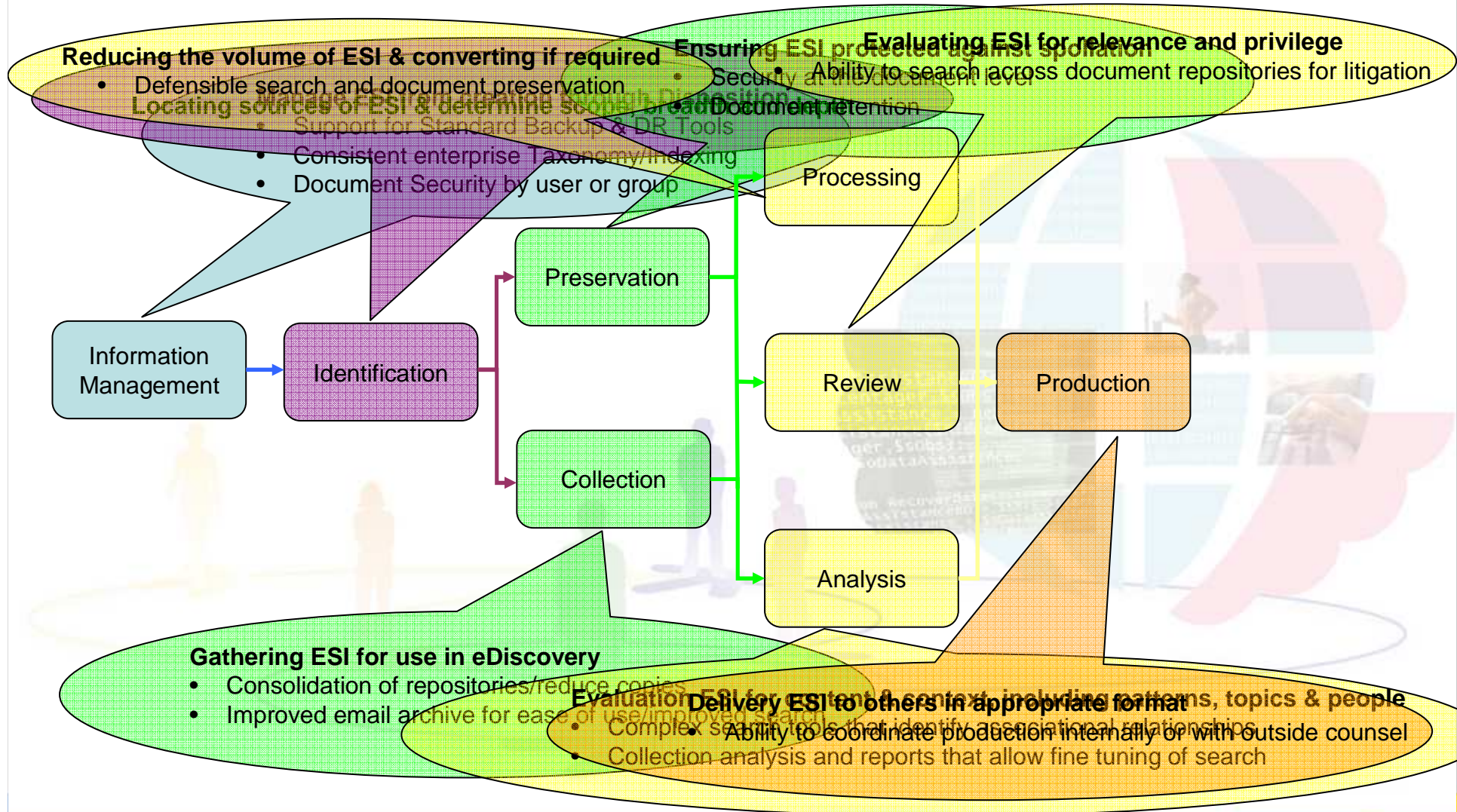
Electronic Discovery (eDiscovery)

- The process of identifying, locating, securing and producing electronically stored information (ESI) and materials, typically in response to litigation





Electronic Discovery Reference Model





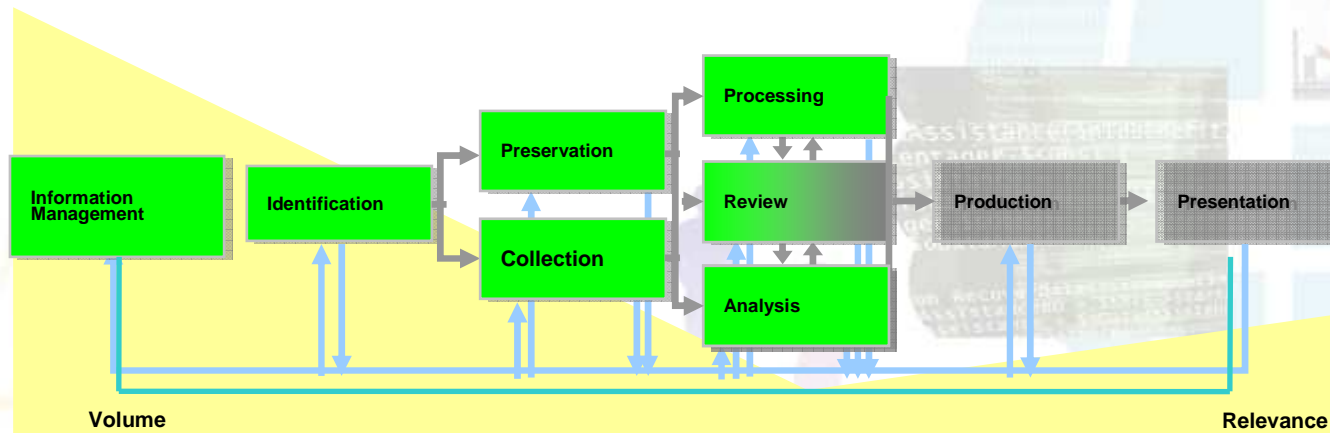
IBM and EDRM

IBM Content Collector
Proactive and reactive
Collection of email, files,
forensic data

eDiscovery Manager
Bulk search,
preservation

eDiscovery Analyzer
First pass review

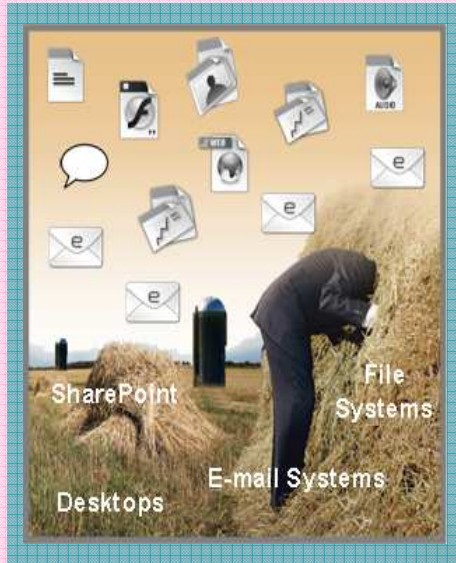
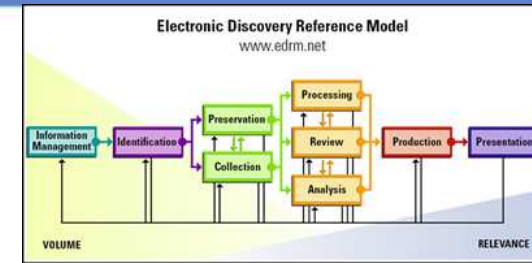
Concordance, etc
Review, production,
presentation



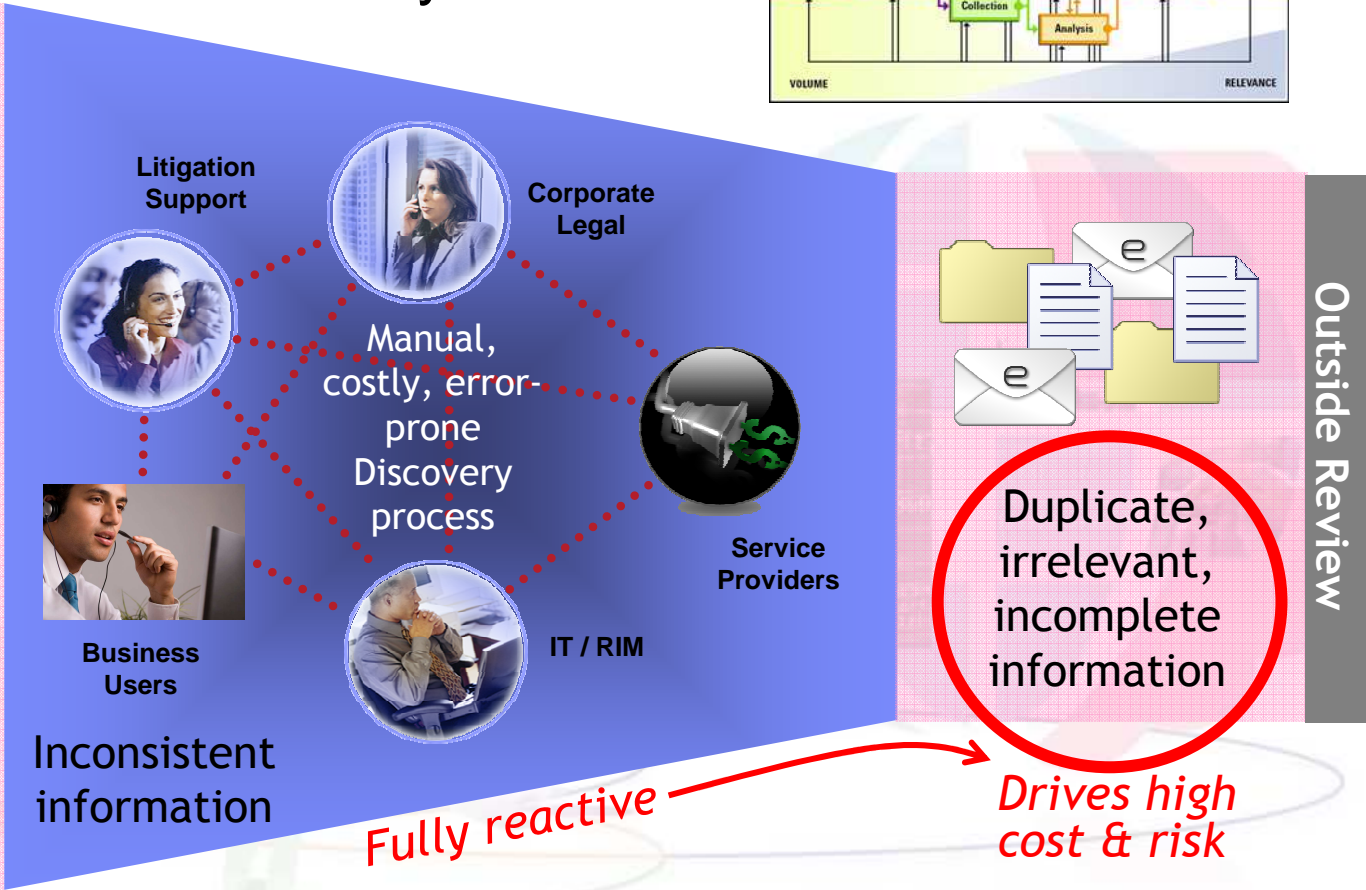
IBM Content Assessment
Analyze information in the wild
in order to decide what to preserve



What Happens Today with Traditional, Reactive eDiscovery?



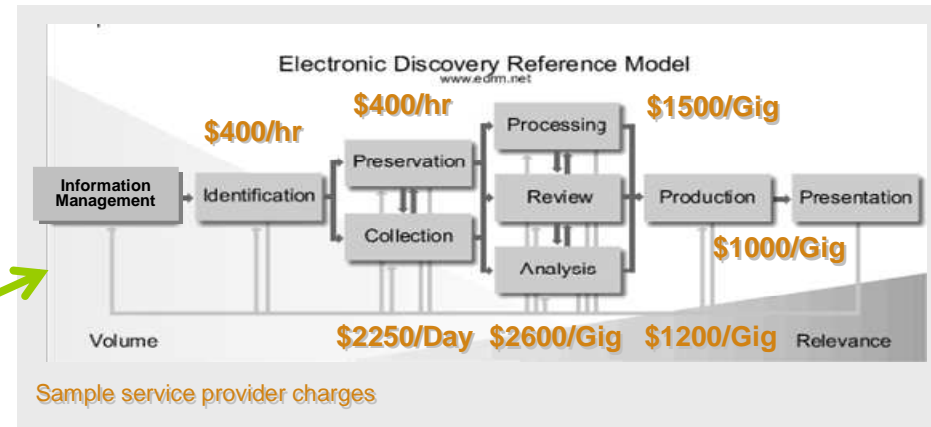
Disorganized, dispersed, lost or overly retained information





Excessive eDiscovery Costs and Risks

- Lack of proactive ESI management creates large volumes of irrelevant or expired information
- Extraneous information translates directly to unneeded eDiscovery cost
- Simple retention solutions like fixed email deletion windows are insufficient



“Information retention programs will be high priorities for companies in all industries...use systematic active policy and strategy for content archiving to reduce legal data processing and attorney review costs by up to a third”

- Gartner Project Planning & Budgeting 2008-11

Case Study: DuPont

- Findings for legal discovery for 9 key cases:
 - Total #pages reviewed: 75,450,000
 - Total #pages responsive: 11,040,000
 - Total %pages past retention period: 50%
 - **Unnecessary review fees: \$11,961,000 USD** at review cost of between 20-80 cents/page
- Does not include other impacts of producing expired information

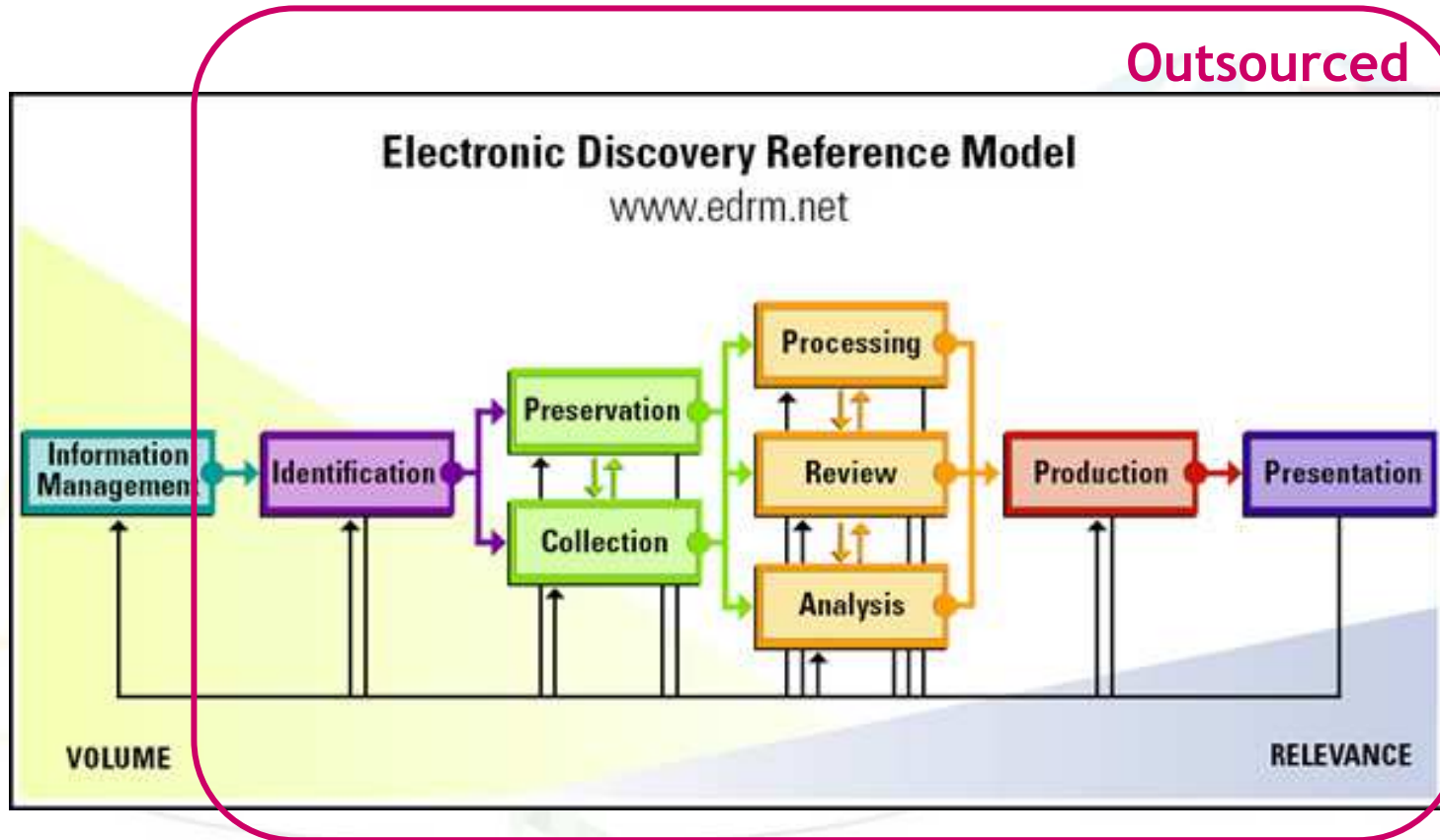




What is “Early Case Assessment?”

- Early case assessment is a hot marketing theme in the e-Discovery space
- In principle, the term refers to the process of evaluating the case accurately before a lot of money and time has been spent in conducting discovery
- In practice, a huge proportion of the expense of litigating a case is spent in the discovery process – early case assessment allows some level of analysis to be done before this money is invested
- eDiscovery Analytics Value Proposition in Early Case Assessment
 - Use it to identify and model potential search terms
 - Use it to identify key and previously unidentified custodians
 - Use it to identify key documents and timelines
 - Enables ability to estimate document volumes

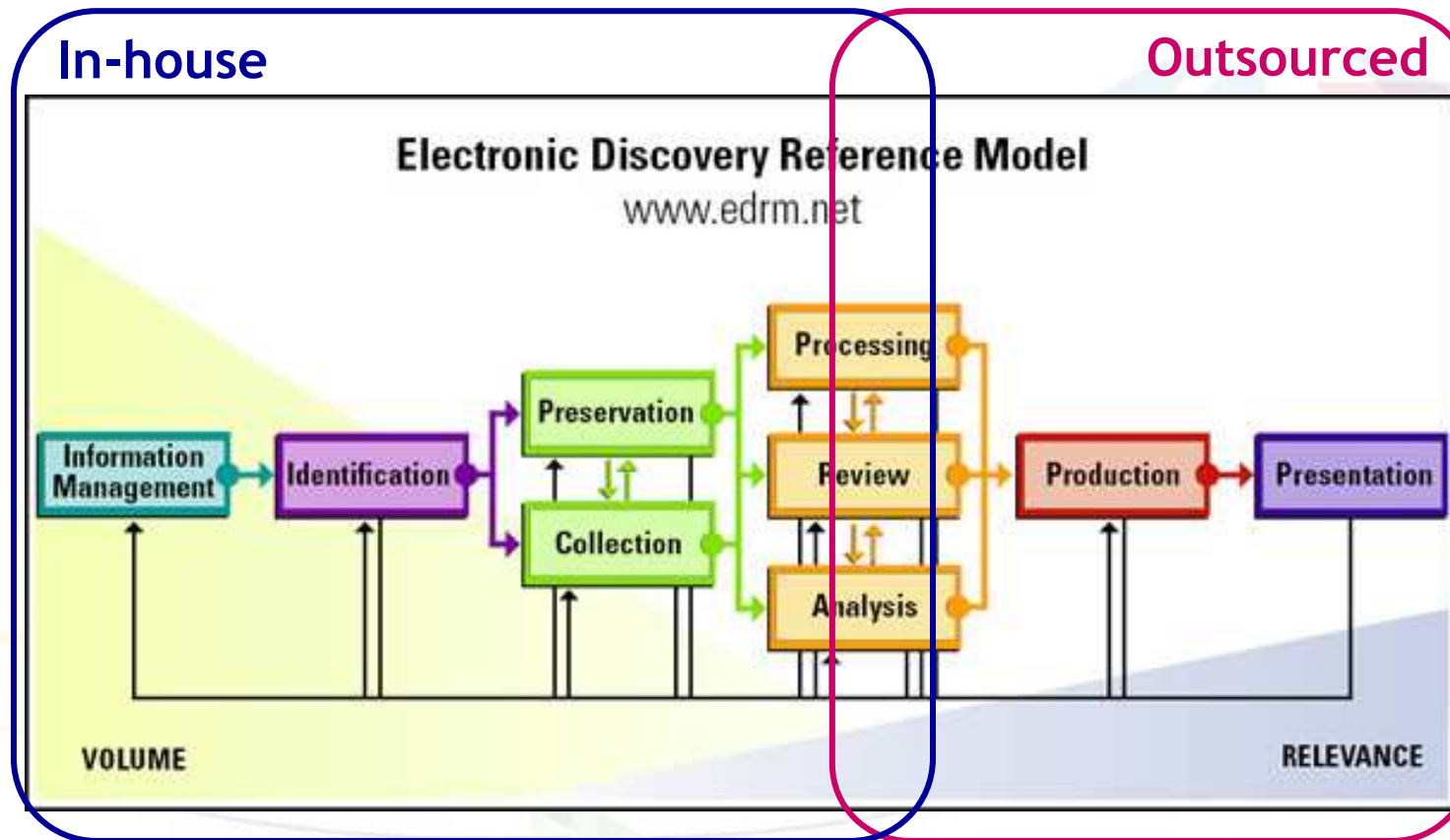
Today eDiscovery is Mostly Outsourced Increased Costs and Increased Risks





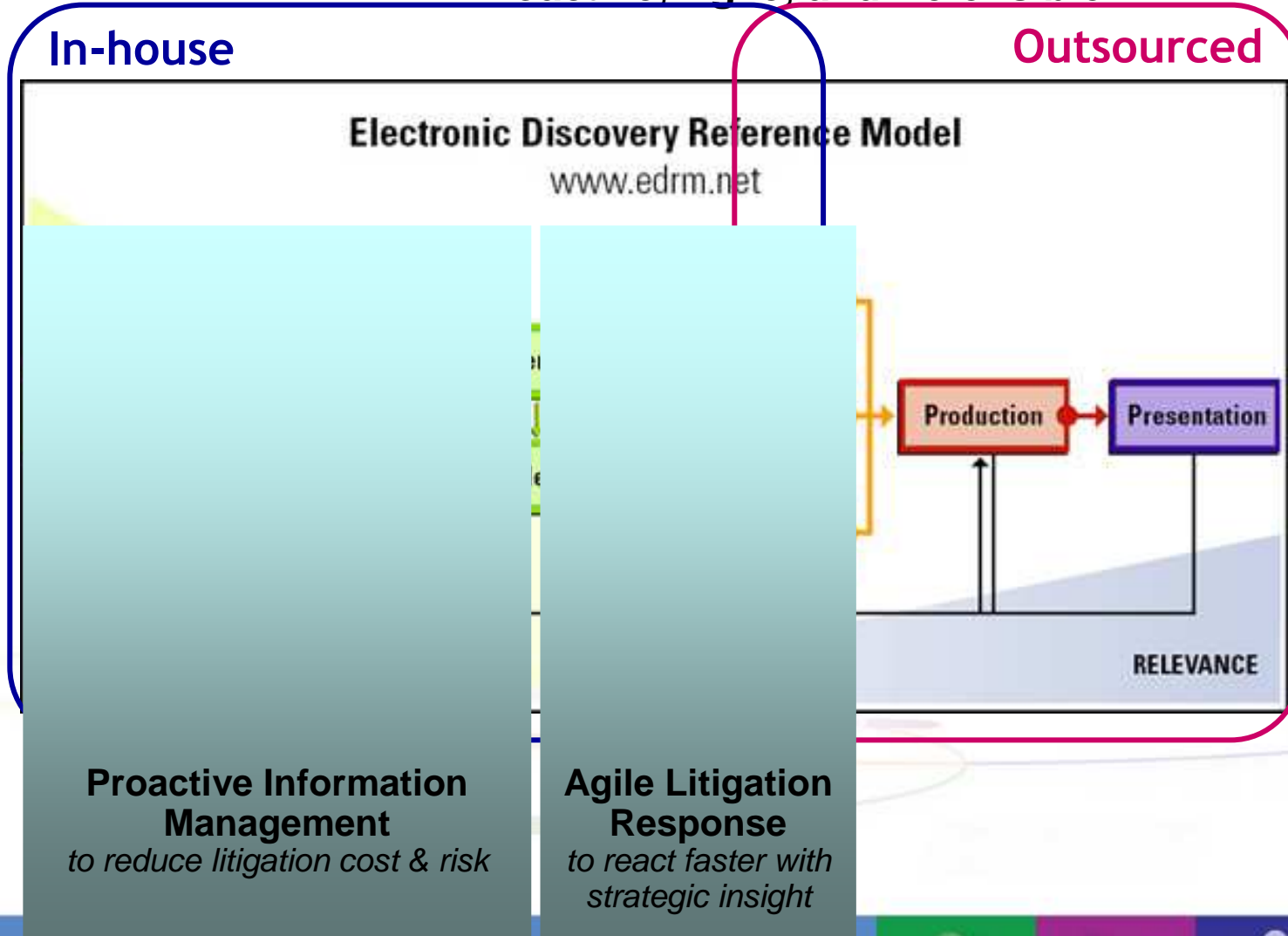
Best Practice Hybrid Approach to eDiscovery

Proactive, Agile, and Defensible



Best Practice Hybrid Approach to eDiscovery

Proactive, Agile, and Defensible



Best Practice Hybrid Approach to eDiscovery

Proactive, Agile, and Defensible

In-house

Outsourced

Electronic Discovery Reference Model

www.edrm.net

Defensible Process
operational benefits & reduced risk

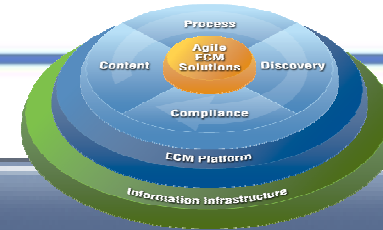
RELEVANCE

Proactive Information Management
to reduce litigation cost & risk

Agile Litigation Response
to react faster with strategic insight



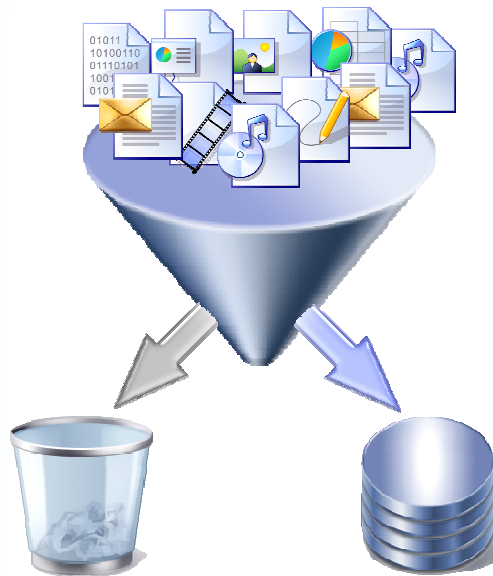
Gain Control with Agile ECM



Gain Control

Know

- Discover, Analyze, Assess and Decide
- Collect or Decommission



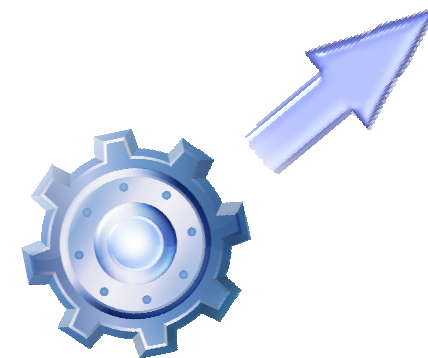
Manage

- Organize and Manage
- Activate, Process and Use
- Archive and Retain
- Trust and Comply

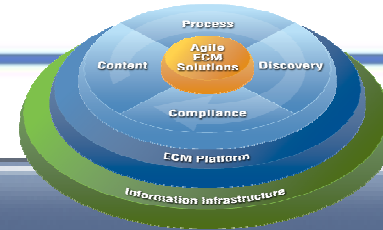


Leverage

- Continuously Optimize
- Analyze, Decide and Predict



Gain Control with ECM eDiscovery



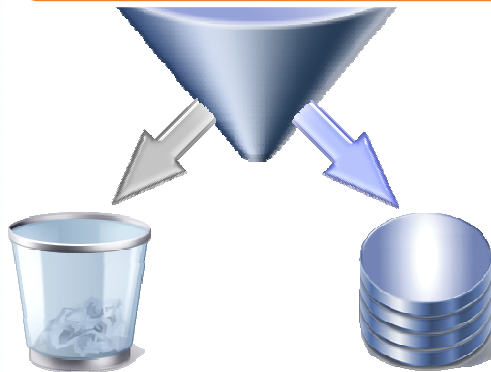
Gain Control

Know

- Discover, Analyze, Assess and Decide
- Collect or Decommission

1

Assess, Decide and Collect
InfoSphere Content Assessment



Manage

- Organize and Manage
- Activate, Process and Use
- Archive and Retain
- Trust and Comply

2

Manage eDiscovery Responses
InfoSphere eDiscovery Manager (eDM)



Leverage

- Continuously Optimize
- Analyze, Decide and Predict

3

Early Case Assess to build
Winning Case Strategies
InfoSphere eDiscovery Analyzer (eDA)



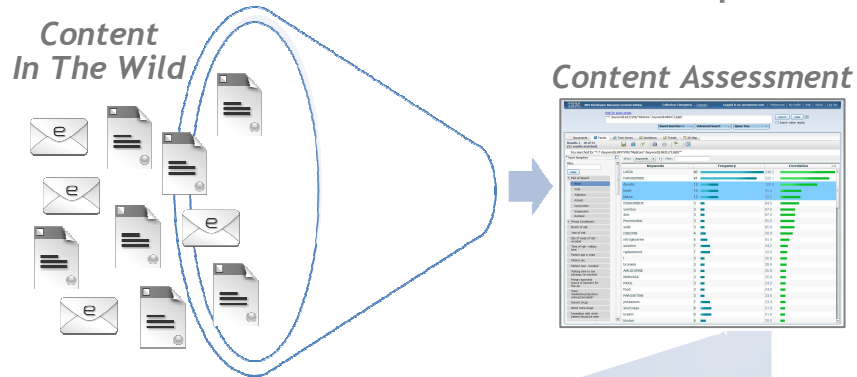


1

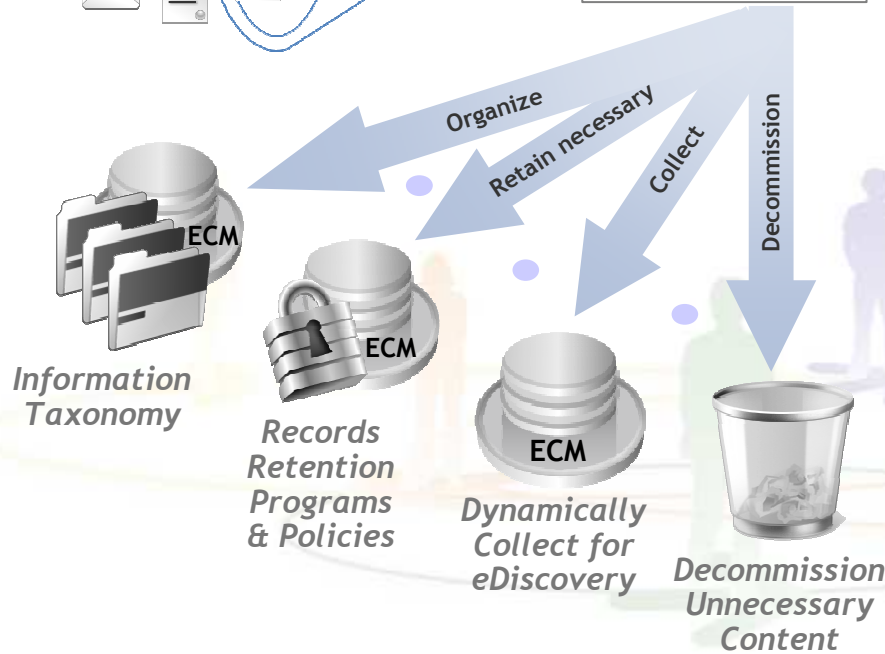
Know: Content Assessment



IBM InfoSphere Content Assessment



...I need to dynamically collect electronically stored information (ESI) by knowing what I have, sorting out the case relevant information and bring under hold management and case analysis..



- Content Assessment enables content-based decision making for:
 - Decommissioning for cost savings – selected content or entire sources
 - **Dynamic collection for eDiscovery**
 - Ongoing proactive information governance
 - Taxonomy bootstrapping
 - Activating content

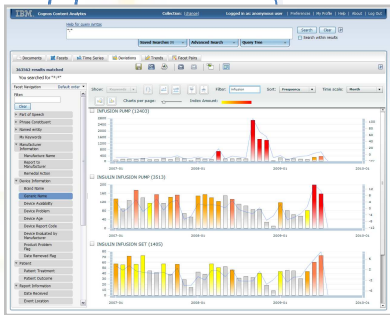




Advancing Policy Based ESI Collection New!

Content Analysis and Assessment

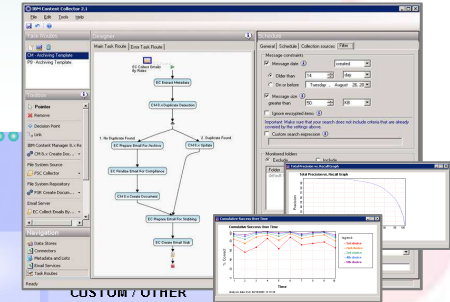
- Configure for “in place” exploration and analysis of desired content sources
- Interactively search, mine and explore information to understand relationships and value to business
- Make decisions, build classification models and collection policies for initial cleanup and ongoing management
- Dynamically collect electronically stored information (ESI) to handle eDiscovery requests



InfoSphere Content Assessment

Continuous Collection and Policy Management

- Setup and manage policies / rules for **Sources**, **Tasks** and **Targets** including de-duplication, records declaration, classification and ongoing monitoring
- Configure and customize **Task Routing** including meta data enhancement, 3rd party applications, advanced context classification and more



Source Connectors

- MESSAGING / EMAIL**
 - Microsoft Exchange / PST
 - Lotus Domino / NSF
 - Lotus iNotes / LotusLive
 - Instant Messaging (3rd Party)
- FILE SYSTEMS**
 - Microsoft File System
 - Desktops (via Tivoli CDF)
 - Data ONTAP (NetApp / IBM N series)
 - Other File Systems
- OTHER**
 - Optim Archive Files New!
 - SAP Archive Files *
 - Microsoft SharePoint
 - Lotus Quickr *
 - IBM Content Integrator
 - CMIS
 - Tivoli Storage Manager
- CUSTOM SOURCES**
 - API Support with SDK

Task Routing Engine

Task Connectors

- INTERNAL TASKS**
 - Copy / Move / Delete / Stub
 - De-Duplication
 - Rules-Based Classification
- EXTERNAL TASKS**
 - Records Declaration
 - Advanced Context Classification
 - Rendering (3rd Party)
- CUSTOM TASKS**
 - API Support / Partner Apps
 - Meta Data Enhancement / Cleansing
 - Encryption
 - System Lookups
 - ILOG Rules Engine (or 3rd Party)
 - Policy Management

CUSTOM / OTHER

- Microsoft File System



InfoSphere Content Collector & InfoSphere Classification Module



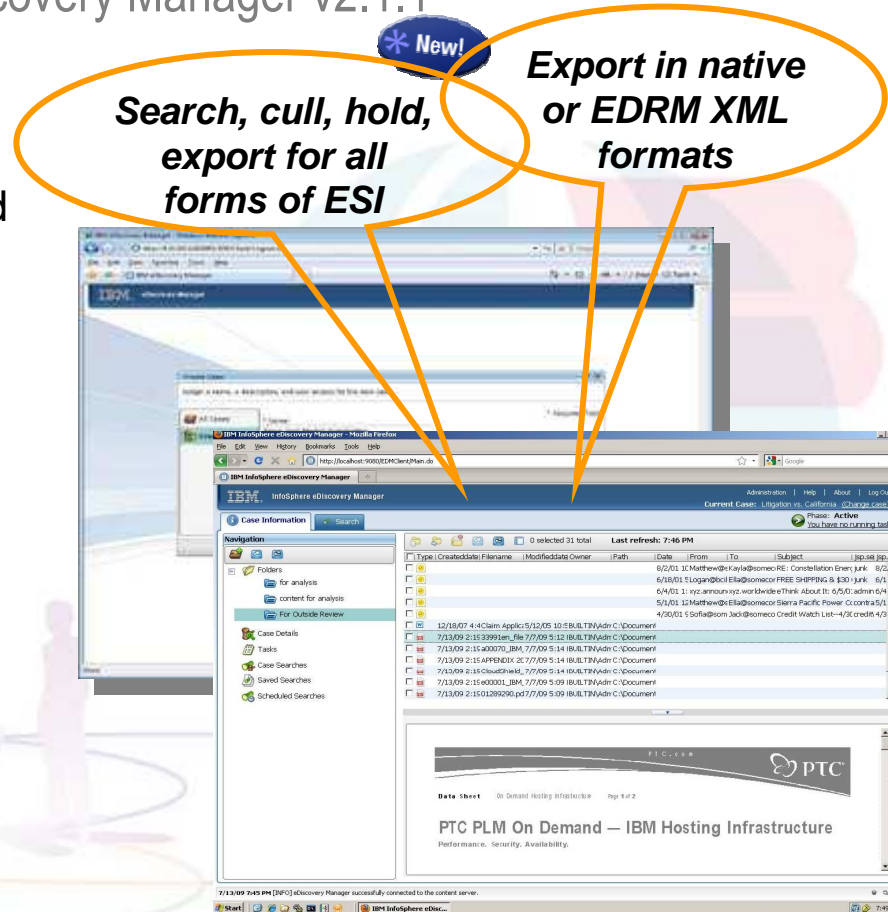


2

Manage : High Volume Smart Culling, Hold, Export

IBM InfoSphere eDiscovery Manager v2.1.1

- **IBM InfoSphere eDiscovery Manager** – the cornerstone of eDiscovery response
- Tool for authorized IT users to help respond to discovery requests from legal
- Provides key functions for IT response to eDiscovery:
 - Creating cases, assigning users to cases
 - Managing ESI in place for multiple cases
 - Searching and culling case relevant ESI
 - Holding and locking down the result set
 - Preview for relevancy
 - Built-in change audit tracking, reporting authenticity and chain of custody
 - Exporting result set in native or EDRM XML format for detailed attorney review
- Built on and integrated with ECM platform to leverage proactive ESI management



IBM InfoSphere eDiscovery Manager – search, cull, hold, export for all ESI types.





3

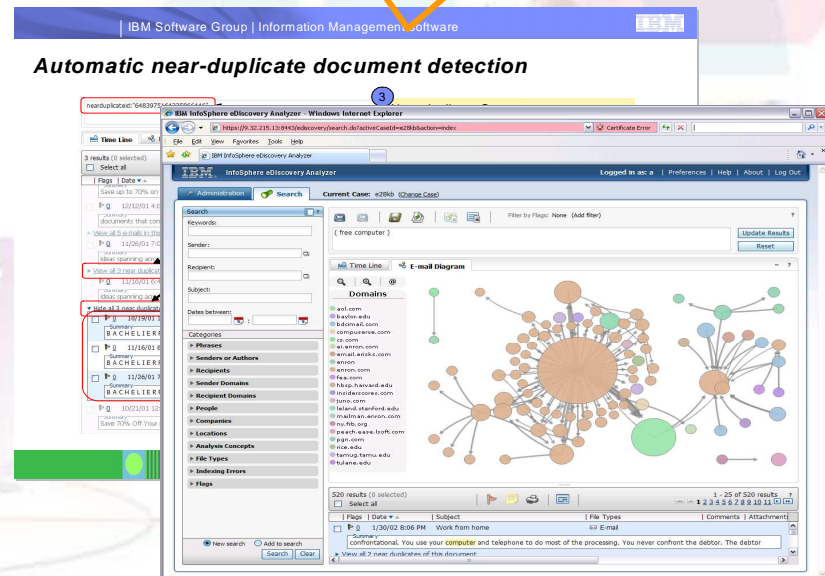
Leverage : Information Insight for Agile eDiscovery

IBM InfoSphere eDiscovery Analyzer v2.1.1

- **IBM InfoSphere eDiscovery Analyzer** – early case assessment & smart culling to increase agility and lower review cost
- Tool for lawyers, legal analysts and supporting specialists to:
 - Quickly reduce case matter to a much smaller and relevant data set
 - Gain early insight into a case, understand key facts, locate key pieces of evidence, and form early case strategies
 - Flag, organize, near-dedup and prioritize documents for review
 - Gain a clear view of available case matter to effectively negotiate queries and term lists during Meet and Confer
 - Preserve chain of custody to ensure security, auditability, and defensibility
- Integrates with and analyzes cases created by eDiscovery Manager

New!

Advanced Early Case Assessment – network diagram, near-dedup, active links, facets, saved searches..



IBM InfoSphere eDiscovery Analyzer – Early Case Assessment, Conceptual Analysis for all ESI types

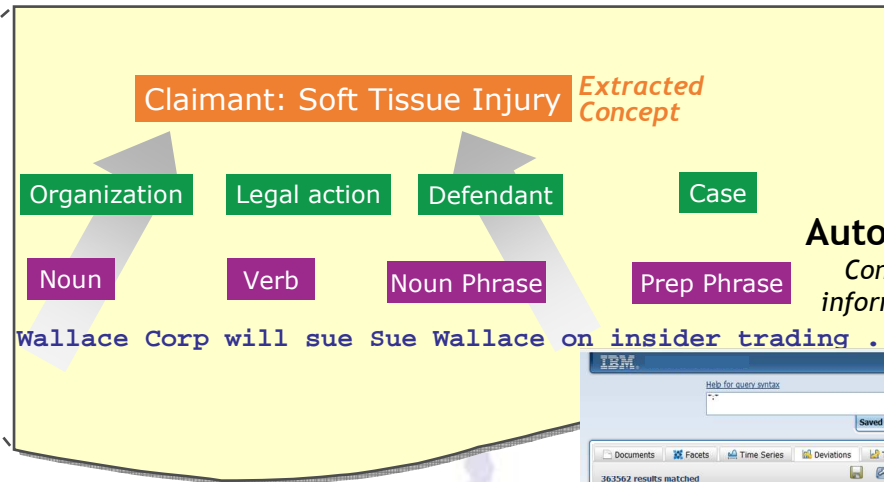


Analytics Making It All Practical to Assist You

Surfacing the Essence of Voluminous Content



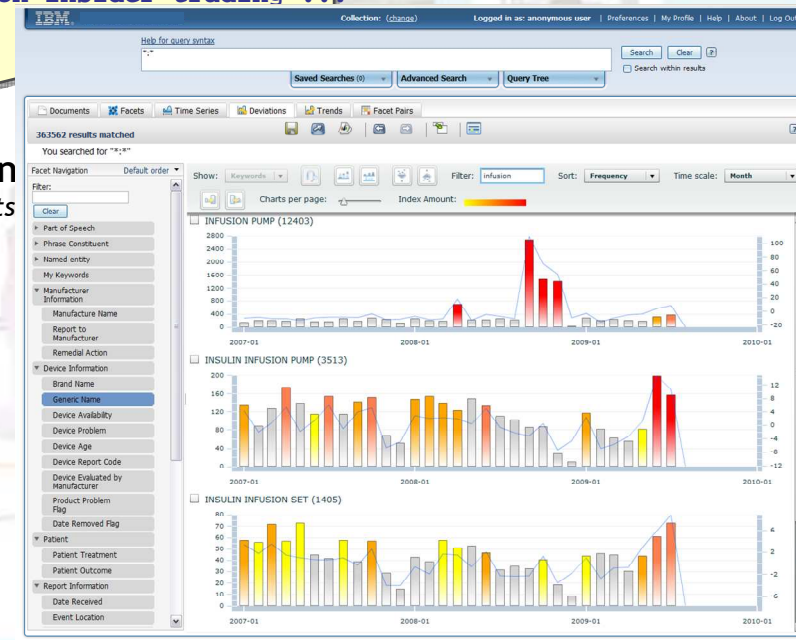
Source Info
(ECM, File, Web, DBMS, ...)



Analyzed Documents
with identified concepts

Automatic Visualization

Concepts and tagged source information are visualized in UI



Based on UIMA, the open, industry-standard architecture for text analysis pioneered by IBM and now an OASIS standard and Apache open-source project



Analytics disambiguates text An example sentence out of documents



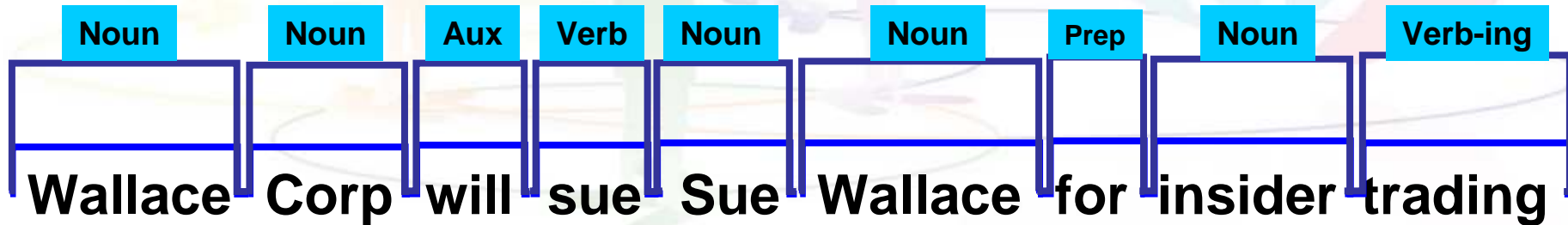
Analytics disambiguates text
Takes each word of the sentences out of a document



Wallace Corp will sue Sue Wallace for insider trading

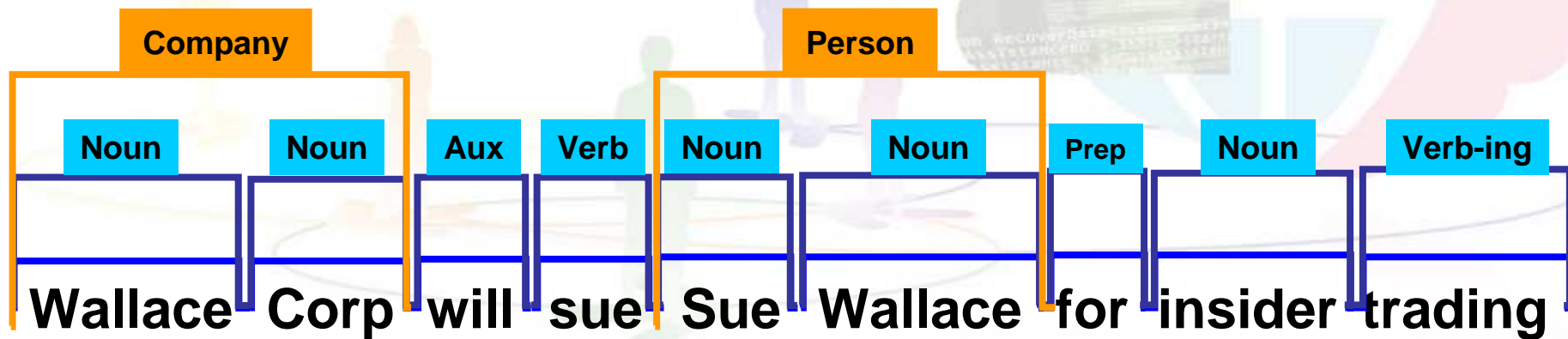


Analytics disambiguates text Finds out what the part of speech each word has

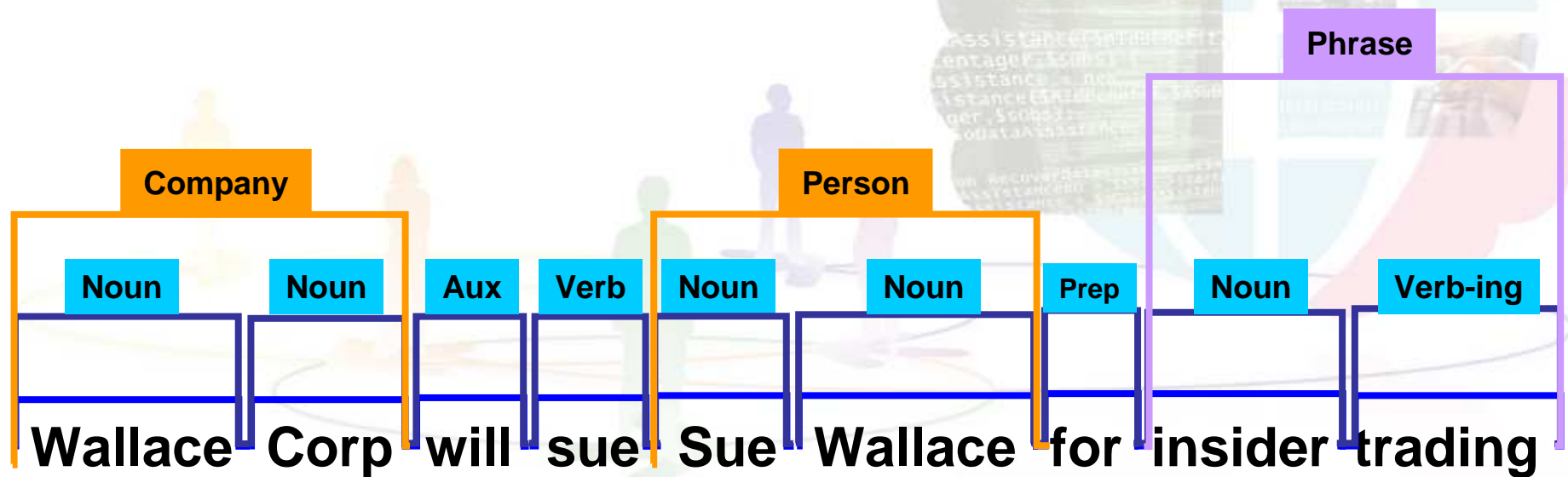


Analytics disambiguates text

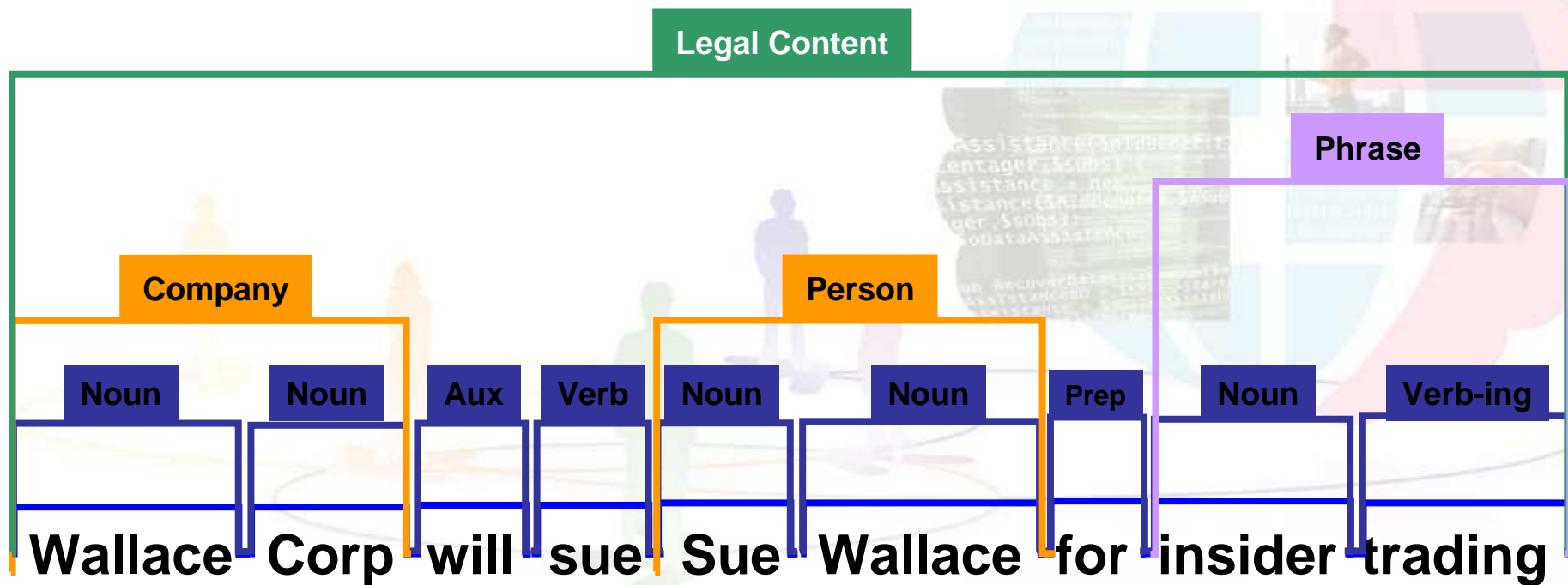
Identifies common entities like person, organizations and locations



Analytics disambiguates text Identifies key phrases



Analytics disambiguates text
Finds if a document has content that matches one of the Analysis Concepts like “Legal Content”





Analytics Empowering Legal Decision Making

- Custodian identification
- Who is talking to whom when, missing custodians
- Volume, complexity
- First pass review
- Near duplicate detection
- Non-responsive e.g. spam, flag them all, take out from the pile
- Iterative search refinement, drill down
- Modeling searches, search terms
- Render the searches you run in boolean logic, negotiate, justify to court

The screenshot displays the IBM InfoSphere eDiscovery Analyzer interface. The top navigation bar includes 'Administration', 'Search', and 'Current Case: Litigation vs. California'. A search query is entered: `(/phrase/"attached file" AND (/phrase/"confidentiality agreement" AND (/legal_concepts/"Confidential Communication" AND (/freetext/"email" AND "*"))))`. The interface features a search sidebar with filters for 'Phrases', 'Senders or Authors', 'Recipients', 'Sender Domains', 'Recipient Domains', 'People', 'Companies', and 'Locations'. The main area shows an 'E-mail Diagram' with a central node 'ella' connected to various other nodes like 'jane', 'kevin', 'jill', 'stacy', 'james', 'brandon', 'anthony', 'nancy', 'nita', 'jack', 'samantha', 'bigabigal', and 'kevin'. Below the diagram, a table of search results is visible, with the first entry dated 3/12/01 8:00 AM and containing a summary about a Confidentiality Agreement.





Increased Defensibility - Justifying What is Done

- Track what is done to separate wheat from the chaff
- Demonstrate how searches, search terms are determined
- What searches, what categories, volume to expect
- How effective they are, yield, result counts
- Get the most out of 26(f) conference and suggested Sedona Cooperation meetings

Query Set Report Print

Case: Litigation vs. California Date: July 17, 2009 06:24:08.733 EDT Report By: lawyer

Queries	Result Count	Singleton Count
[Combined Query]	3761	3761
(/recipients/"ella@somecompany.com")	2934	2385
(/recipients/"samantha@somecompany.com")	98	83
(/recipients/"lauren@somecompany.com")	68	51
(/recipients/"addison@somecompany.com")	60	43
(/recipients/"nancy@someplace.com")	58	43
(/recipients/"sofia@acompany.com")	1	1
(sell OR buy OR trade OR purchase)	922	344
(account)	223	66
(financials)	27	0
(statement)	159	31

[Expand All](#) [Collapse All](#)

- [Combined Query]
- (/recipients/"ella@somecompany.com")
- ▾ (/recipients/"samantha@somecompany.com")
 - Result Count: 98
 - Singleton Count: 83
 - ▾ Category: Sender Domains
 - somecompany.com - 96
 - lotusspace.com - 2
 - Category: Phrases
 - Category: Companies
 - Category: Locations
 - Category: Recipients



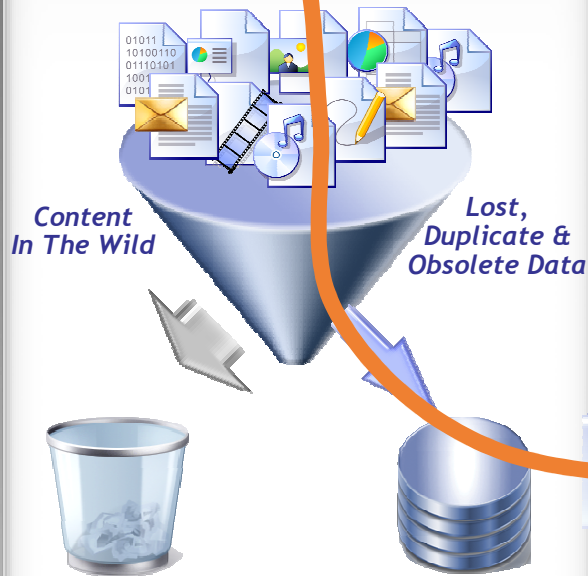
Optimizing eDiscovery End-to-End to Reduce Cost & Risk



Gain Control

Know

Bloated Production Systems with Inefficient Storage



- InfoSphere Content Assessment
- InfoSphere Content Collector (for Email, for Files, for SharePoint)

Manage

- IBM Smart Archive
- InfoSphere Classification Module
- InfoSphere Content Collector (for Email, for Files, for SharePoint)
- FileNet Business Process Manager
- FileNet Content Manager
- InfoSphere Enterprise Records
- InfoSphere eDiscovery Manager



Leverage

- Reduced review costs
- Improved agility and better case strategy
- Reduced risk with increased defensibility



- InfoSphere Master Content
- Cognos Content Analytics
- InfoSphere eDiscovery Analyzer

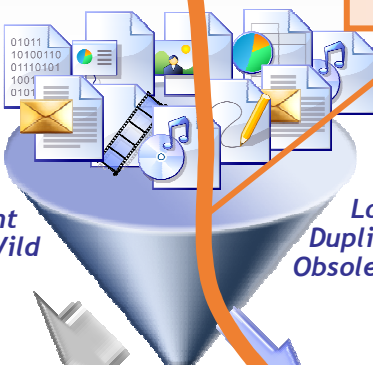
Optimizing eDiscovery End-to-End to Reduce Cost & Risk



Gain Control

Know

Bloated Production System with Inefficient Storage



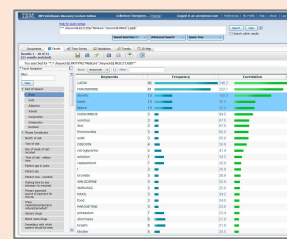
Content In The Wild

Lost, Duplicate & Obsolete Data

- InfoSphere Content Assessment
- InfoSphere Content Collector (for Email, for Files, for SharePoint)

Content Assessment enables dynamic, on-demand evaluation and collection of content in the wild

Manage



- IBM Smart Archive
- InfoSphere Classification
- InfoSphere Content Manager (for Email, for Files)
- FileNet Business Platform
- FileNet Content Manager
- InfoSphere Enterprise Records
- InfoSphere eDiscovery Manager

Leverage

- Reduced review costs
- Improved agility and better case strategy
- Reduced risk with increased defensibility

- InfoSphere Master Content
- Cognos Content Analytics
- InfoSphere eDiscovery Analyzer

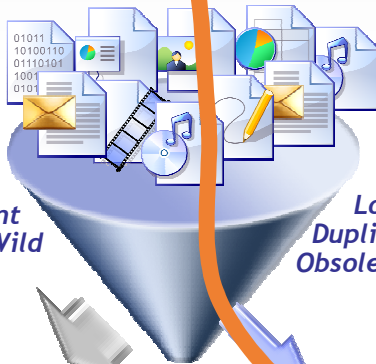
Optimizing eDiscovery End-to-End to Reduce Cost & Risk



Gain Control

Know

Bloated Production Systems with Inefficient Storage



Content In The Wild

Loss of Duplicate Obsolete Data



- InfoSphere Content Assessment
- InfoSphere Content Collector (for Email, for Files, for SharePoint)

Manage

Content Collector enables proactive and reactive collection of critical information into secure, audit-tracked ECM repository

- InfoSphere Enterprise Records
- InfoSphere eDiscovery Manager



Leverage

- Reduced review costs
- Improved agility and better case strategy
- Reduced risk with increased defensibility



- InfoSphere Master Content
- Cognos Content Analytics
- InfoSphere eDiscovery Analyzer

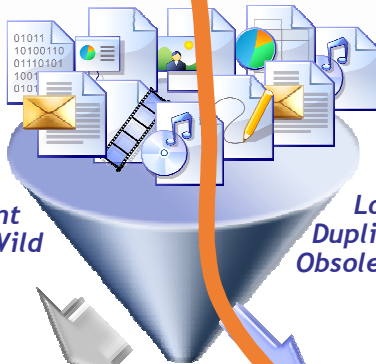
Optimizing eDiscovery End-to-End to Reduce Cost & Risk



Gain Control

Know

Bloated Production Systems with Inefficient Storage



Content In The Wild

*Low Duplication
Obsolete Data*



- InfoSphere Content Assessment
- InfoSphere Content Collector (for Email, for Files, for SharePoint)

Manage

Enterprise Records and Classification Module enable retention and disposition management of collected content

- InfoSphere Enterprise Records
- InfoSphere eDiscovery Manager



Leverage

- Reduced review costs
- Improved agility and better case strategy
- Reduced risk with increased defensibility



- InfoSphere Master Content
- Cognos Content Analytics
- InfoSphere eDiscovery Analyzer

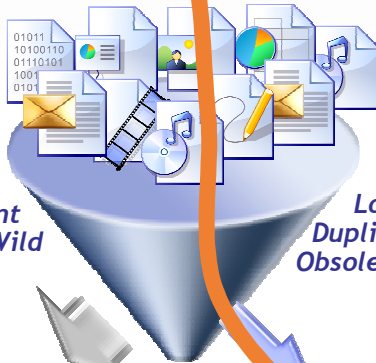
Optimizing eDiscovery End-to-End to Reduce Cost & Risk



Gain Control

Know

Bloated Production Systems with Inefficient Storage



Content In The Wild

Lost, Duplicate & Obsolete Data



- InfoSphere Content Assessment
- InfoSphere Content Collector (for Email, for Files, for SharePoint)

Manage

eDiscovery Manager provides central tool to manage legal holds, case-based ESI foldering, initial search and culling, and ESI export



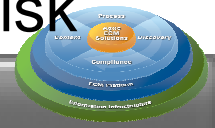
Leverage

- Reduced review costs
- Improved agility and better case strategy
- Reduced risk with increased defensibility



- InfoSphere Master Content
- Cognos Content Analytics
- InfoSphere eDiscovery Analyzer

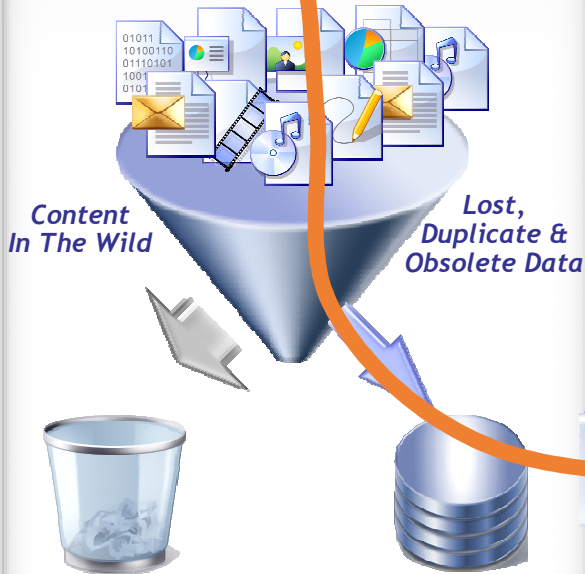
Optimizing eDiscovery End-to-End to Reduce Cost & Risk



Gain Control

Know

Bloated Production Systems with Inefficient Storage

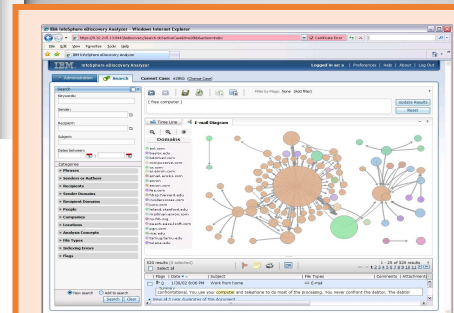


- InfoSphere Content Assessment
- InfoSphere Content Collector (for Email, for Files, for SharePoint)

Manage

eDiscovery Analyzer provides business user (legal) interface with rich analytics for early case assessment, smart culling decisions, and investigations

• InfoSphere eDiscovery Manager



better case strategy
Reduced risk with increased defensibility



- InfoSphere Master Content
- Cognos Content Analytics
- InfoSphere eDiscovery Analyzer