

# Business Optimisation with Content Discovery: Leveraging Classification, Content Analytics, and Search for Better Business Outcomes

**Brendan Reid**  
**Global ECM Solutions Specialist**



**DATA MANAGEMENT**

*DATA SERVERS, DATA WAREHOUSE  
AND DATA ARCHIVING (PRINCETON)*



**ENTERPRISE CONTENT  
MANAGEMENT**

*INCLUDING FILENET PRODUCT PORTFOLIO*



**INFORMATION INTEGRATION &  
MASTER DATA MANAGEMENT**

*PREVIOUSLY KNOWN AS ASCENTIAL,  
DWL, SRD & TRIGO.*



**BUSINESS INTELLIGENCE &  
PERFORMANCE MANAGEMENT**

*COGNOS*

# Content Discovery for Business Optimisation

***How much time are your employees wasting trying to find or organize content?***

## **Cohasset Associates study:**

*Typical 1000-person organization can spend **\$170K+/month** in lost productivity to manual records declaration*

## **Large crime intelligence agency:**

*More than **50% of analysts time** spent looking for content rather than analyzing it*

***How well are you leveraging content in your business decisions?***

**80% of information in today's enterprises is unstructured content.**

*Yet traditional decision support tools like BI work only on the 20% of the information that's structured.*

**What insights are you missing as a result?**

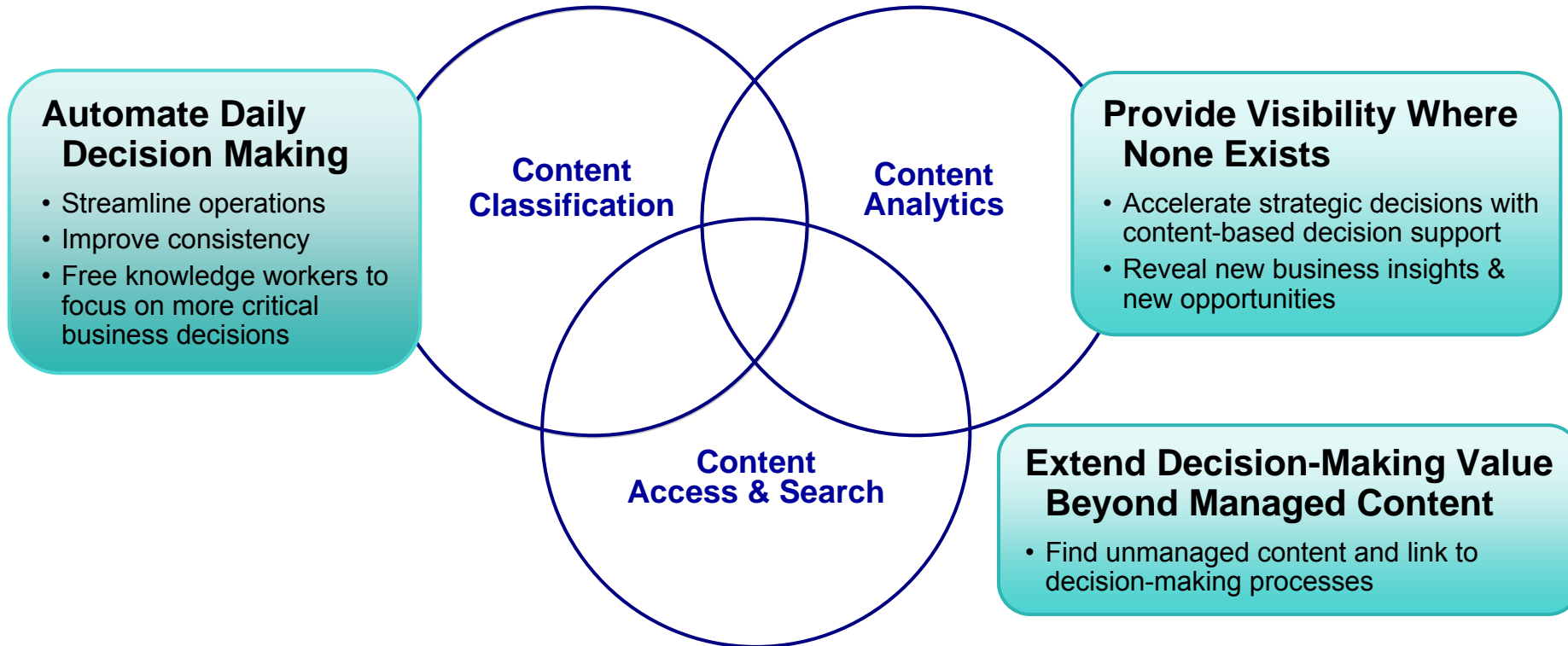
***What if you could automate content-based decisions and expose new business insights from content, automatically?***



# Content Discovery for Business Optimisation

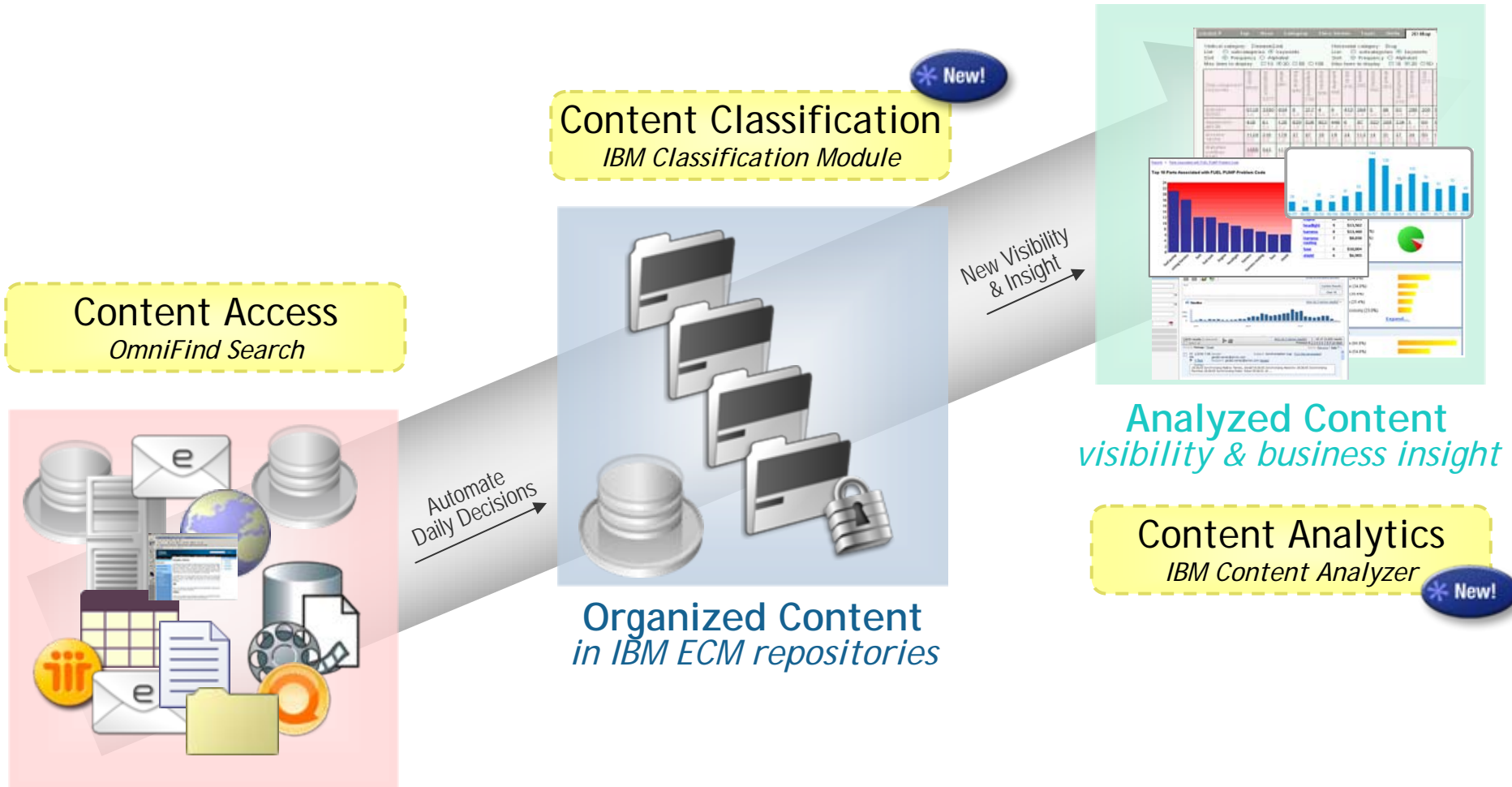
- **You can!** Content Discovery products from IBM extend your ECM environment to optimise business decision making

## ECM Content Discovery Capabilities



# Content Discovery Portfolio

Driving visibility and business optimisation from enterprise content



Disorganized or Lost Content





# Agenda

- Introduction
- ➔ **Automating Daily Decisions: IBM Classification Module**
- **Extending Beyond Managed Content: OmniFind search**
- **Providing Visibility Where None Exists: IBM Content Analyzer**
- **Wrap-up**



# Content Discovery Portfolio

Driving visibility and business optimization from enterprise content

## Classification: Automate Daily Decision Making

- Streamline operations
- Improve consistency
- Free knowledge workers to focus on more critical business decisions

## Content Classification

*IBM Classification Module*

**New!**



**Organized Content**  
*in IBM ECM repositories*

New Visibility & Insight



**Analyzed Content**  
*visibility & business insight*

## Content Analytics

*IBM Content Analyzer*

**New!**

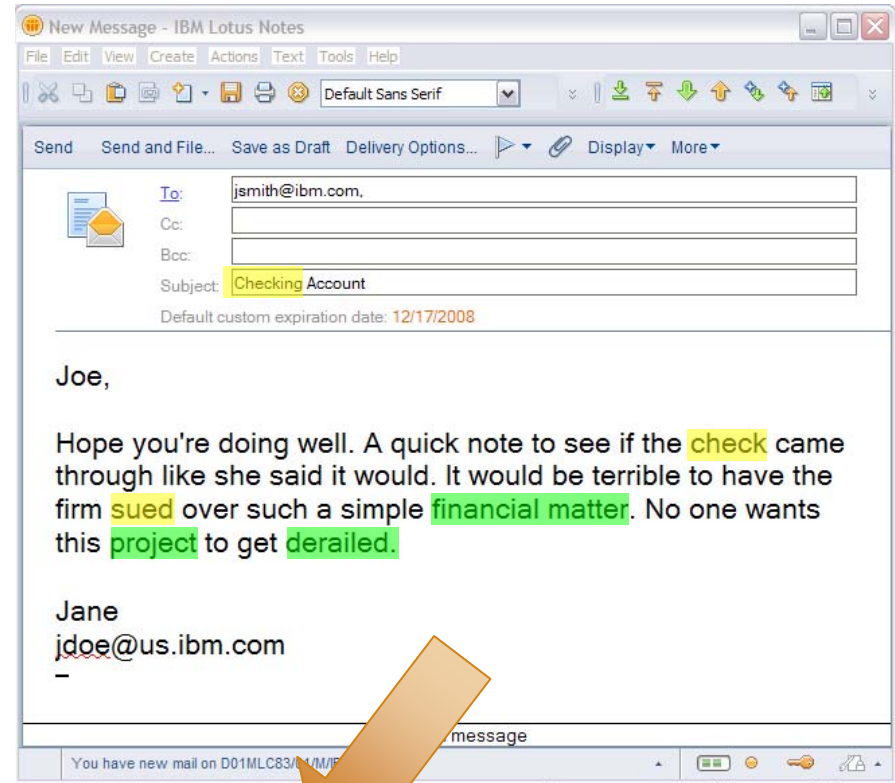
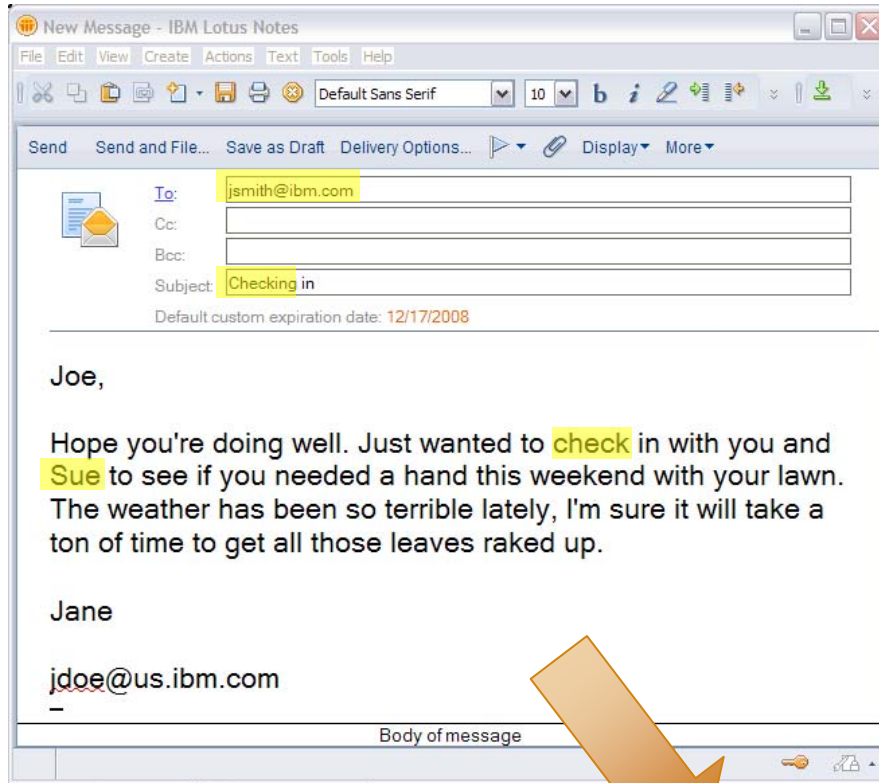


**Disorganized  
or Lost Content**



# What is Classification?

Leveraging the full content context to make the right decision



# Why Classification? *Velocity and Volume of Daily Decisions*

- **A typical knowledge worker might**

- receive 50 emails
- send 20 emails
- take part in 20 instant message conversations
- create 5 new documents every day

- **80% of information in the enterprise is unstructured**

- 6.6 trillion business emails sent annually
- A typical enterprise might archive 500,000 emails per day



# Manual Classification is a Path to Failure

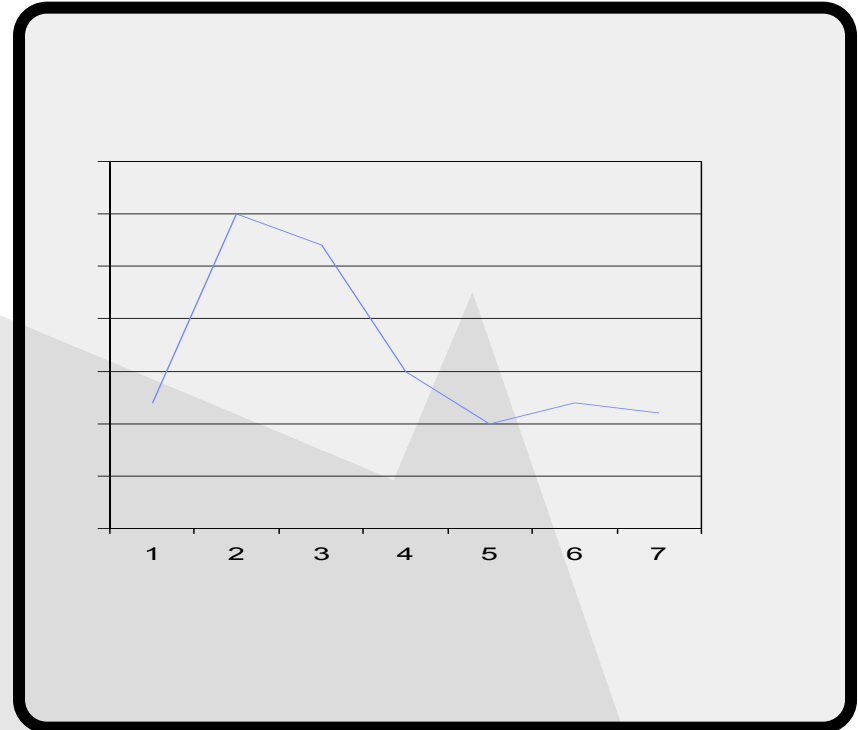
- **National Archives and Records Administration**

- Electronic Records Management initiative focused on user driven records declaration
- 6+ month study
- Significant user drop-off after training period

- **End users frequently outright refuse to categorise content**

- **Silos full of existing content abound resulting in large backlogs in addition to new content**

- **Manual classification and an emphasis on “user training” an outdated practice**



# The Solution: IBM Classification Module

## Expanding the ECM platform with a robust, adaptive classification service

The screenshot displays the IBM Classification Module interface. At the top, it shows the IBM logo and the title 'Classification Module / Classification Review Tool'. The interface includes a navigation menu on the left with options like 'View: Single Document', 'View: Document List', and 'Add New Document'. The main area shows a document titled 'underlined name' with a file icon and details like '(338768-1747926.txt, 1 KB)'. Below this, there are buttons for 'Previous', 'Next', 'Apply Classification', 'Delete Document', 'Re-run Classification', and 'Mark as Unknown'. A section titled 'Folders Classification' shows 'Suggested Folders' with a table of folder names and percentages, and 'Selected Folders' with a list of folders. Another section, 'Document Class Classification', shows 'Suggested Document Classes' with a dropdown menu set to 'QA/News Article - 98.45%'. The interface also includes a 'Start Over' button and a 'Reclassify All' button in the left sidebar.

Folder Name	Percentage
QA/Base/Science News	94.79%
QA/Base/News	9.3%
QA/Base/Movies	5.89%
QA/Base/Haaretz	2.59%
QA/Base/DeveloperWorks	1.39%



### Automates Content-Sensitive Decisions

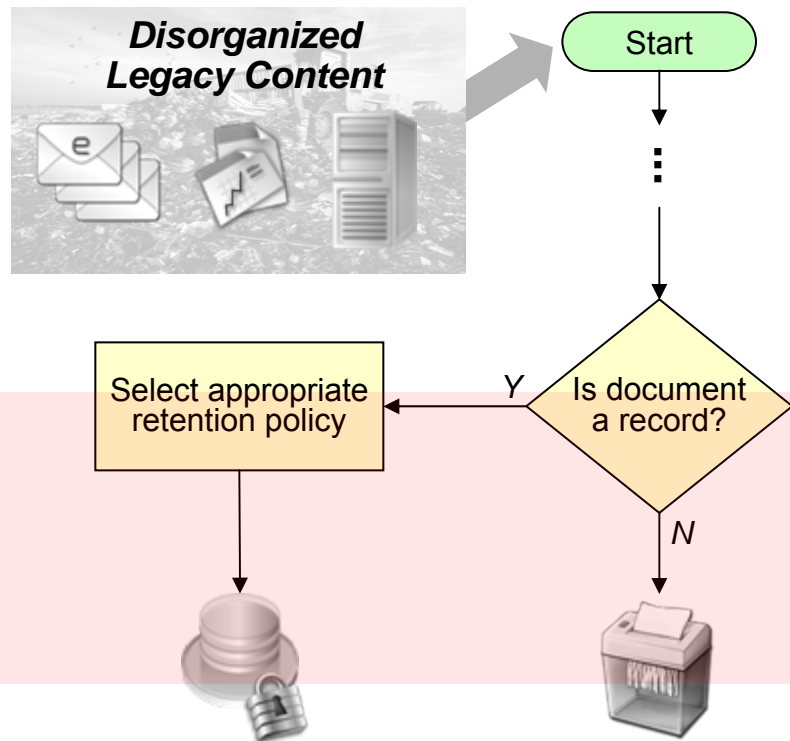
- Automated classification and filtering
- Combines adaptive text understanding with rules
- Acquires domain specificity from your own content
- Unique learning technology for adaptive classification improves accuracy over time
- Integrated with IBM ECM architecture and in applications such as Records Crawler
- Suggests new categories or even seeds an entirely new taxonomy
- 5<sup>th</sup>-generation technology; market proven, scalable platform supporting 12 languages



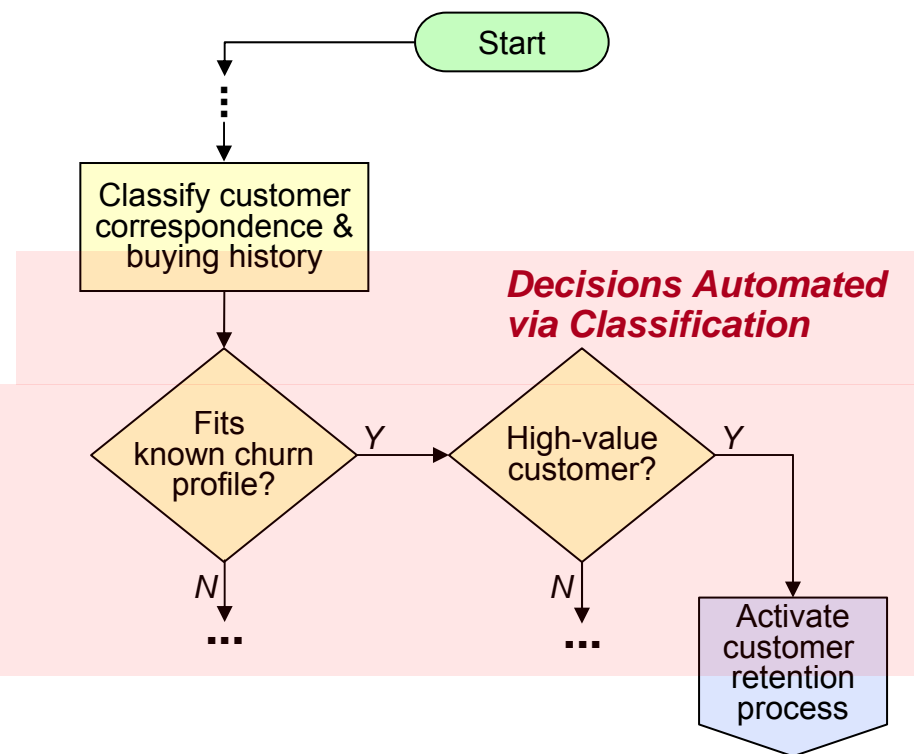
# Classification Module Automates ECM Decision-Making

- Adaptive classification automates content-based decisions in formal and ad-hoc workflows, improving productivity and reducing cost

*Compliance Readiness Process:*



*Customer Support and Retention Management Process:*



# Example: Classification Automates Records Management Decisions

**Disorganised Content**

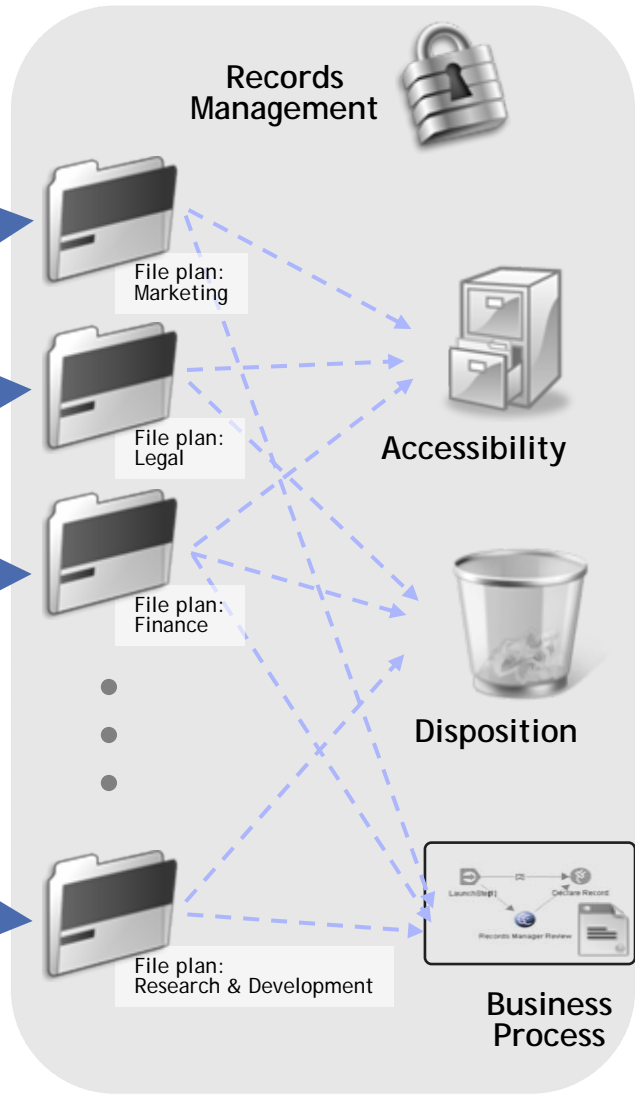
- Existing ECM Repositories
- File Shares
- Basic Content Services
- Email



**IBM Classification Module**



Review & Audit



# Example: IBM Compliance Warehouse for Legal Control

Leveraging IBM End-to-End Capabilities...including classification

**Disorganised**  
Enterprise Content  
Manual Processes  
Paper Storage

**Organise / Control**  
*Automatic Classification*  
*Records Management*  
*Intelligent Content Archiving*

**Optimise**  
*Legal Discovery*  
*Operational Controls*  
*Analytics, BI & Reporting*



**Ingest**  
Tagging for Discovery  
& Process Context

**Enable**  
Streamline Discovery  
& Legal Productivity

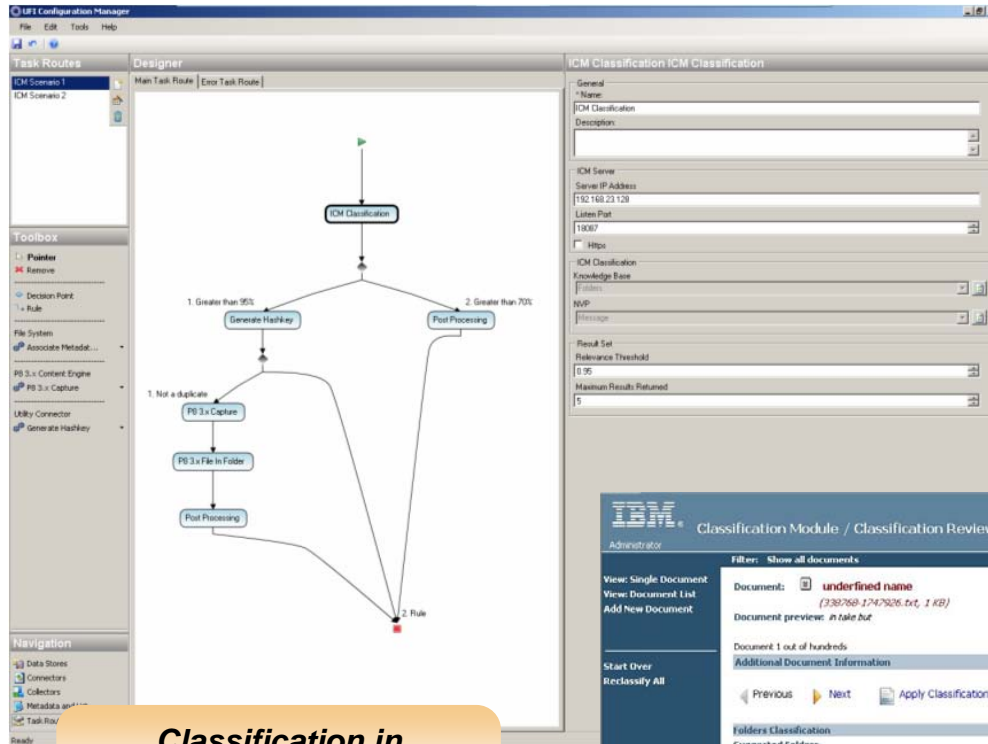


# More Use Cases where Classification Adds Value

- **IBM Classification Module improves ROI by automating decisions in a broad range of ECM-based solutions:**
  - **Loan origination:** better automating routing based on analysis of free-form descriptions and comments
  - **Claims processing:** claim prioritization, process routing, fraud estimation based on descriptions, correspondence, interviews, and more
  - **Correspondence management:** automated routing and handling based on determination of request type
  - **Compliance management:** automated identification & declaration of records from poorly managed sources like SharePoint and file systems
  - **Vertical document management:** accelerated solution deployment and reduced long-term cost with automatic mapping of existing documents into appropriate vertical taxonomy
  - **Legal/regulatory matter management and legal discovery:** faster organization and prioritization of content; adaptive similarity analysis
  - *Similar value-add opportunities exist in process optimisation for insurance, utilities, government-specific solutions, and more*



# IBM Classification Module in ECM



**Reclassification Application and Review Tool for P8**

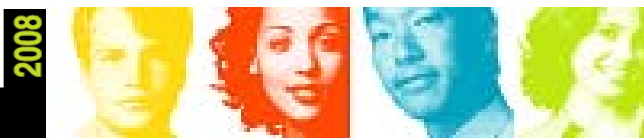
**Classification in Records Crawler – and Symphony in the future**

Suggested Folders	
select QA/Base/Science News	94.79%
select QA/Base/News	9.3%
select QA/Base/Movies	5.89%
select QA/Base/haaretz	2.59%
select QA/Base/DeveloperWorks	1.39%
All folders...	

Suggested Document Classes	
QA/News Article	- 98.45%
All document classes...	

**All powered by underlying reusable, scalable classification service**



# Cloudmark

*Improving service while reducing costs with IBM Classification Module*



## ■ Challenge

- Reduce support center workload and costs while improving customer service

## ■ Solution

- IBM Classification Module helping to classify and auto-respond to customer e-mails and support requests

*"Without the IBM Classification Module, we would have had to double our team to handle our growing customer base, which would have cost hundreds of thousands of dollars each year."*

Kris Politopoulos, Manager,  
Cloudmark Customer Support

## ■ Results

- 100% payback in less than three months
- 400% annualised return on investment
- 15% increase in customer retention
- 50% reduction in number of e-mails received
- 96% faster solution deployment than with competitors' products
- Annual savings of hundreds of thousands of dollars



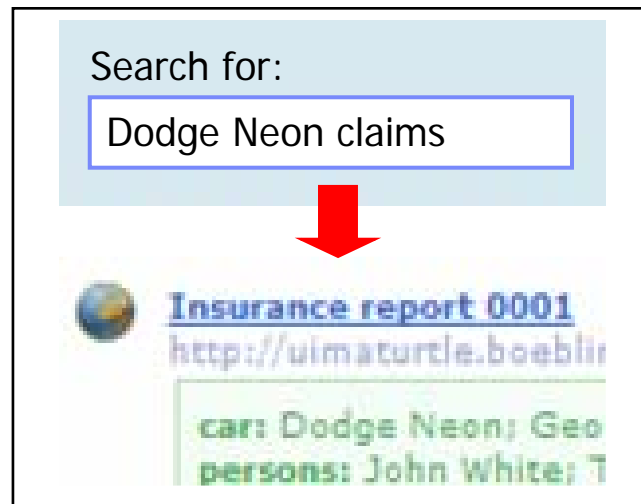
# Agenda

- Introduction
- Automating Daily Decisions: IBM Classification Module
- ➔ ■ **Extending Beyond Managed Content: OmniFind search**
- **Providing Visibility Where None Exists: IBM Content Analyzer**
- Wrap-up



# Business optimisation beyond managed content

- **Search & Content Access extend optimised decision-making to unmanaged content**
  - Extends beyond the confines of managed content to unmanaged content sources
    - Web pages, blogs, wikis, forums – internal or on the public Internet
    - File systems
    - Scattered enterprise repositories – Domino DBs, databases with text fields, ...
    - SharePoint and other “viral” collaboration sources
  - Integrates with classification and content analytics to deliver insight on “found” content
- **And delivers insight with a familiar, effective search-box metaphor**



# Example: Search + Analytics deliver decision-making insight from public discourse

- **Online Media Analysis solution helps accelerate strategic marketing and PR decisions based on unmanaged web and public Internet content**
  - Solution leverages OmniFind Enterprise Edition and IBM Content Analyzer

The screenshot displays a search interface for 'hybrid car' with a search bar and a 'Search' button. Below the search bar, there is a link to 'Add this search to My Searches'. The results are refined by 'Brands Toyota' and sorted by 'Relevance' in 'Descending' order. A 'Discovery Dashboard' is visible on the left, showing a pie chart for 'Tonality' with the following data:

Tonality	Percentage
positive	83.6%
negative	8.4%
neutral	8.0%

Below the tonality chart, 'Hot Terms' are listed with horizontal bars representing their frequency:

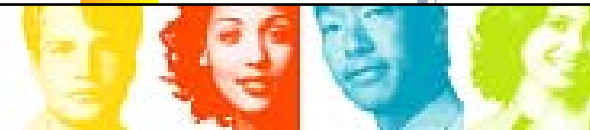
Hot Term	Percentage
prius	54.0%
nissan	34.0%
lexus	30.4%
camry	25.4%

The search results list '2006 Toyota Highlander Hybrid Car Reviews - The Car Connection' with a date of 1/3/07. A red circle highlights a green sentiment icon next to the title. The snippet of the article text is: '2006 Toyota Highlander Hybrid Car Reviews The Car Connection. ...It 's pretty simple to boil Toyota 's Highlander Hybrid down to this It 's the Lexus RX400h 's drivetrain ...Now Toyota has seriously upped the Highlander 's profile by stuffing in the same Hybrid Synergy Drive system that 's ...While the RX400h 's and Highlander 's hybrid system carries the same Hybrid Synergy Drive name as that in Toyota ...In fact the RX400h 's and Highlander 's system is in some ways more advanced than its hybrid cousin. ...' Below this, another result for '2006 Toyota Highlander Hybrid Car Reviews - The Car Connection' is shown with a date of 12/28/06 and a 'positive' sentiment icon.

Search & browse interface

Content auto-classified by sentiment

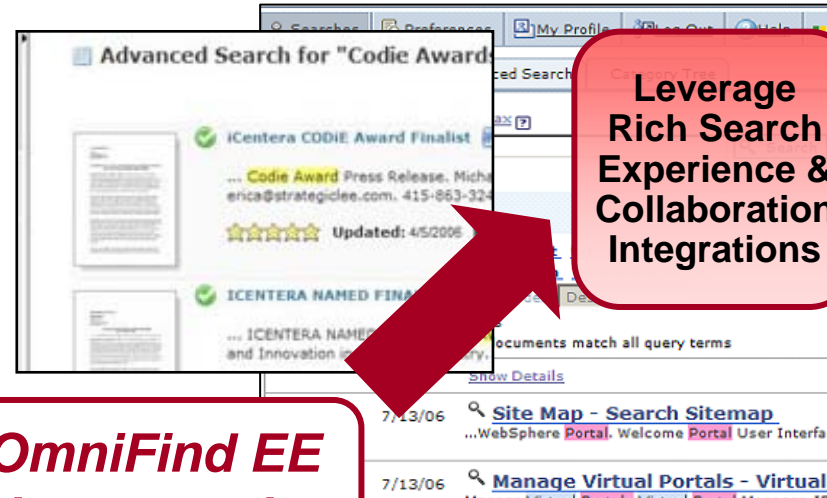
Analytics-powered interactive dashboard



# IBM OmniFind Enterprise Edition also delivers leading content access capabilities for content solutions

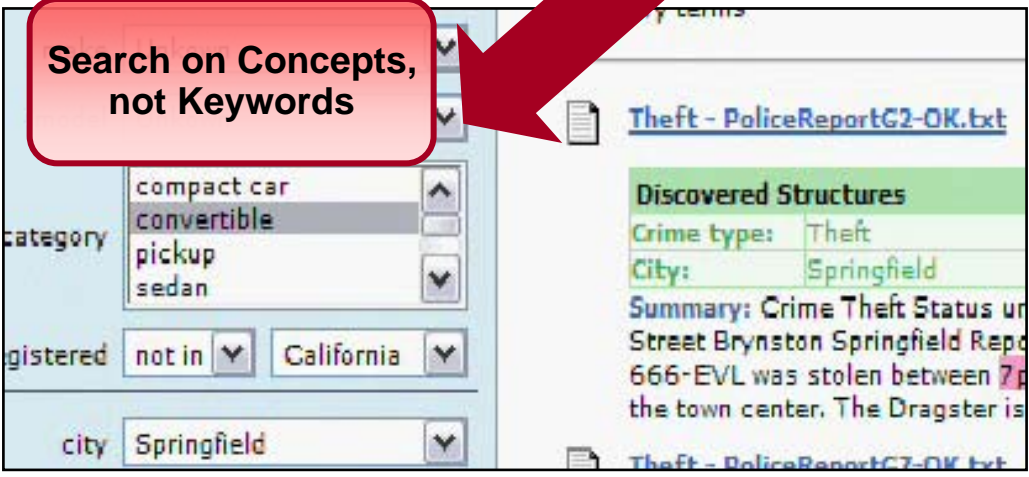


**ECM repository only provides basic search**

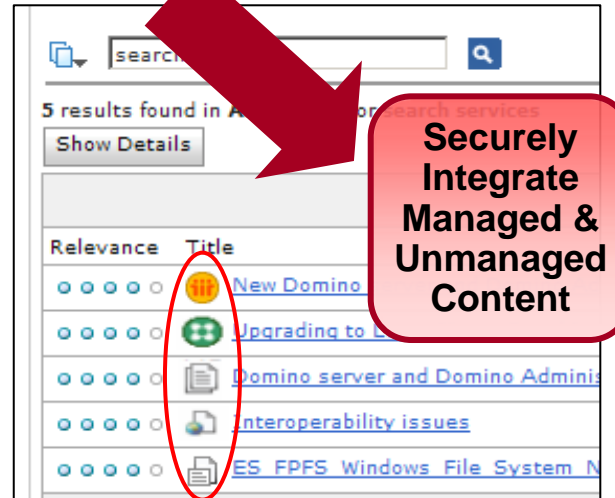


**Leverage Rich Search Experience & Collaboration Integrations**

**But adding OmniFind EE enables much more value**



**Search on Concepts, not Keywords**



**Securely Integrate Managed & Unmanaged Content**

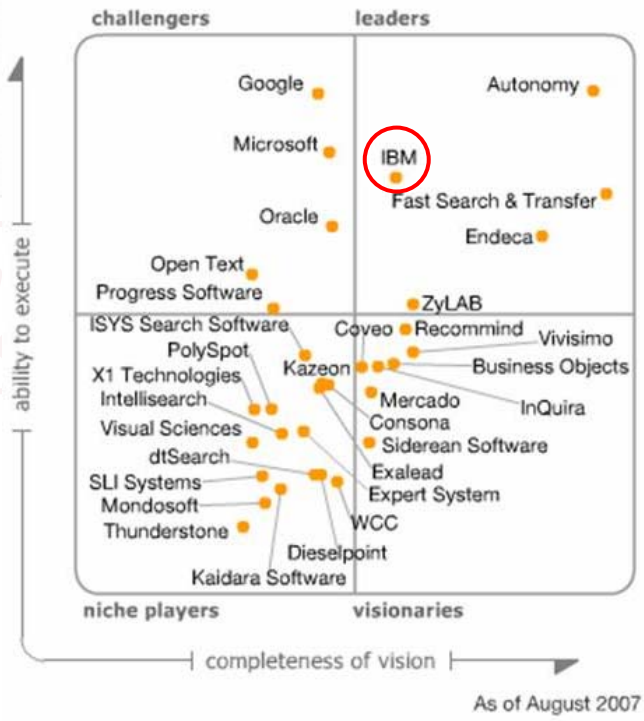


# OmniFind Enterprise Edition also delivers leading content access capabilities for content solutions

## OmniFind Enterprise Edition

- ▶ Recognised as a Leader in the 2007 Gartner Information Access Magic Quadrant
- ▶ Adds robust enterprise search and access capabilities to content solutions
- ▶ Securely integrates content from across the enterprise, even unmanaged content
- ▶ Rich APIs for powering custom UIs and applications
- ▶ Integrates with IBM Classification Module and IBM Content Analyzer

**New v8.5 released in March '08!**



The Magic Quadrant is copyrighted 2007 by Gartner, Inc. and is reused with permission. The Magic Quadrant is a graphical representation of a marketplace at and for a specific time period. It depicts Gartner's analysis of how certain vendors measure against criteria for that marketplace, as defined by Gartner. Gartner does not endorse any vendor, product or service depicted in the Magic Quadrant, and does not advise technology users to select only those vendors placed in the "Leaders" quadrant. The Magic Quadrant is intended solely as a research tool, and is not meant to be a specific guide to action. Gartner disclaims all warranties, express or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

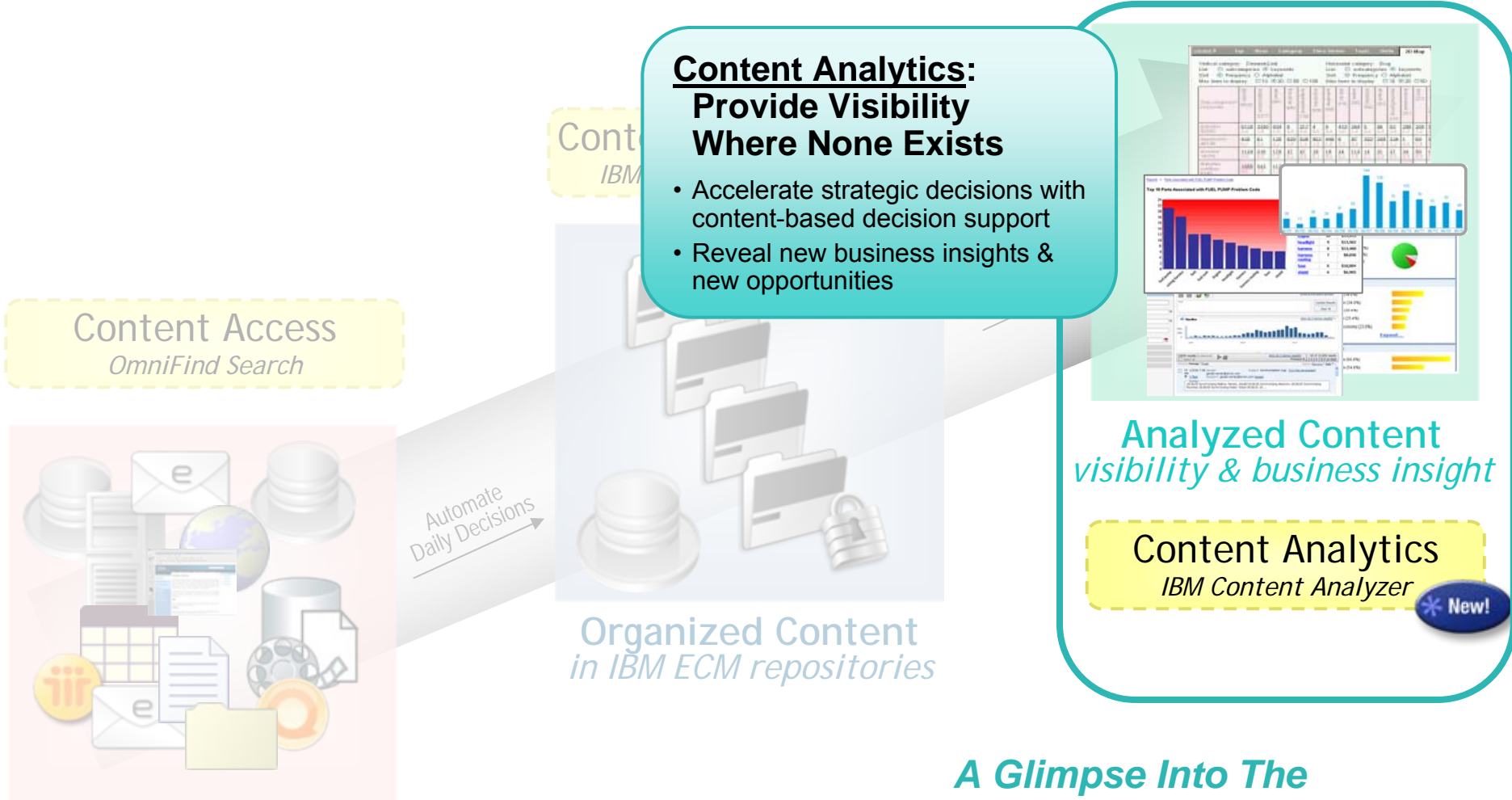
# Agenda

- Introduction
- Automating Daily Decisions: IBM Classification Module
- Extending Beyond Managed Content: OmniFind search
- **Providing Visibility Where None Exists: Content Analytics**
- Wrap-up



# Content Discovery Portfolio

## Driving visibility and business optimisation from enterprise content

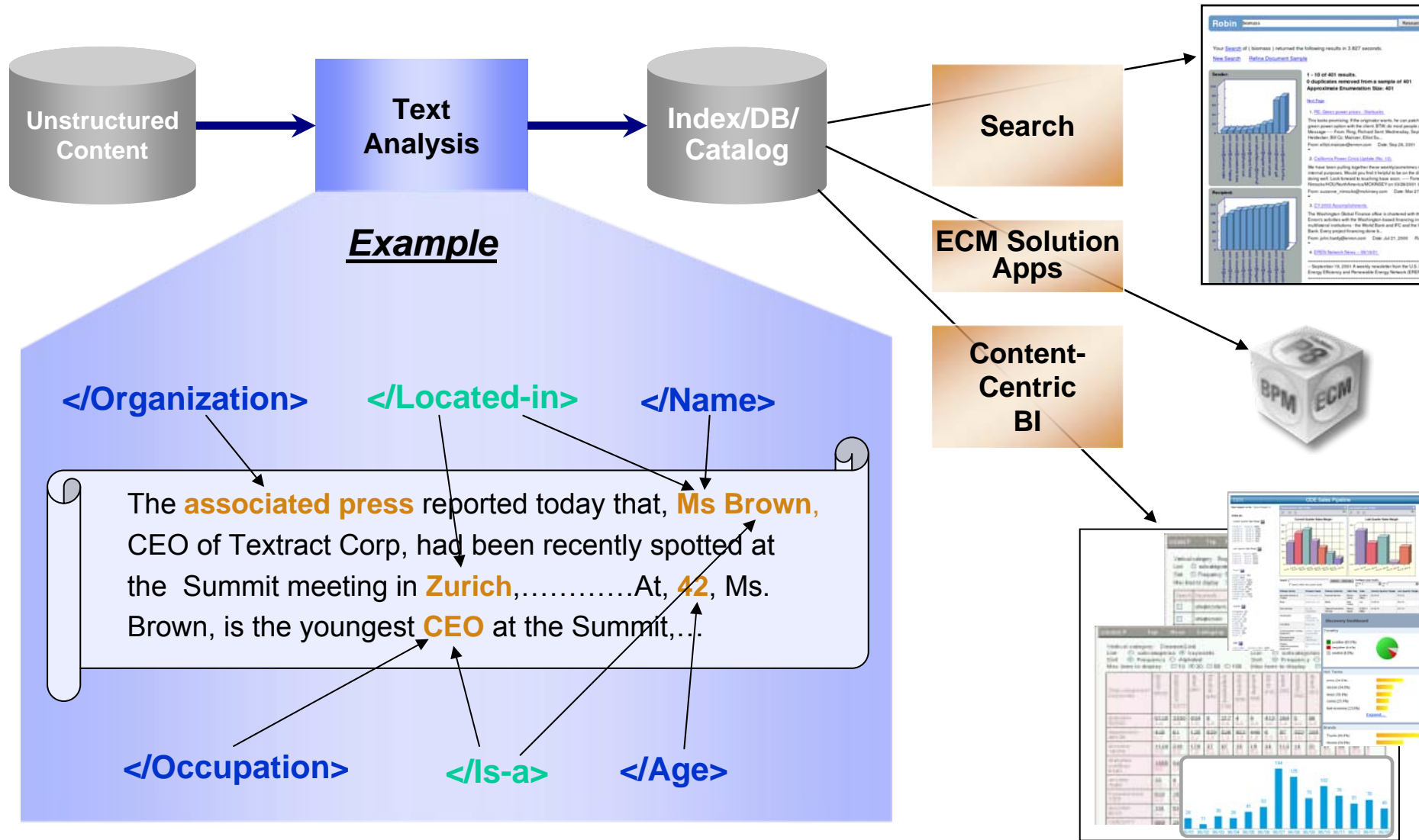


Disorganized or Lost Content

*A Glimpse Into The Future...*



# The Key to Content Visibility: Text Analytics



Powered by UIMA – the open, industry-standard Unstructured Information Management Architecture pioneered by IBM



# What can Content Analytics do for You?

- **Provide visibility into your content**

- Identify aggregate patterns, trends over time, unusual correlations or anomalies
- Power content dashboards
  - To deliver insights from business lines that run on content
  - To identify compliance risk
  - To enhance database-backed BI/EPM systems with insight from content

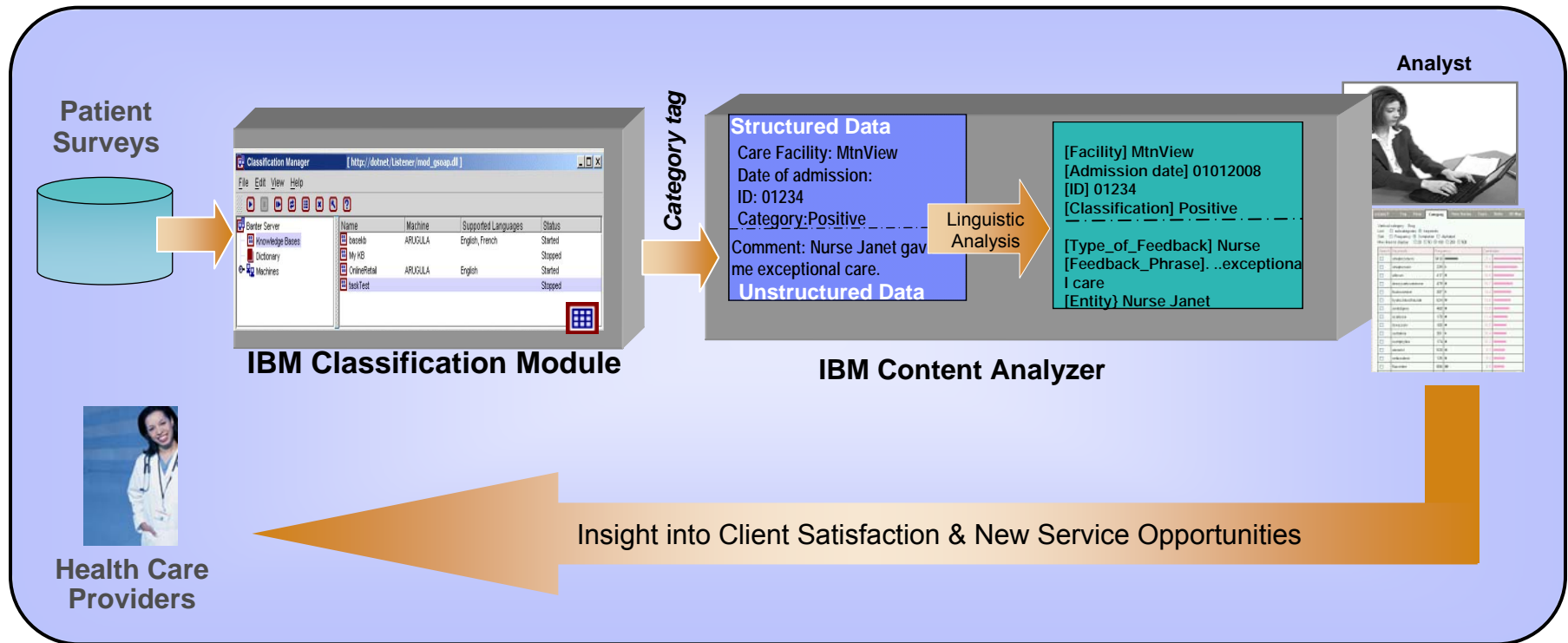
- **Increase the ROI on your ECM investment**

- Extend your content investment: deliver new services that provide real business insight – go beyond content housekeeping
- Enhance your process investment: gain visibility into how content and process interact, driving process optimization
- Accelerate your compliance initiatives: deploy tools to help automate eDiscovery & compliance investigations, and provide risk visibility



# Example: 360° View of the Customer Using Content Analytics to Expose New Service Opportunities

- Content analytics plus classification on customer surveys provides insight into new value-add services, improving profitability & client satisfaction



Healthcare Provider Organisation



# Example: Predictive Content Analytics in Context Compliance Investigations & eDiscovery

- Content analytics provides insight in context of compliance investigation and legal discovery processes, reducing risk and cost of compliance

The screenshot shows a software interface for a compliance investigation. At the top, it says 'Current Case: Cheatham vs. JK Enterprises'. Below this are tabs for 'Case Information', 'Collect', 'Analyze', 'Task Manager', 'Audit', and 'Administration'. The 'Analyze' tab is active.

On the left, there are search filters for 'Keywords', 'Senders', 'Recipients', 'Subject', and 'Between Dates'. Below these are 'Review Flags' including 'Non-Responsive', 'Responsive', 'For Further Review', and 'To be redacted'. There are 'Include' and 'Exclude' buttons at the bottom left.

The main area is divided into several sections:

- Search Filters:** Shows 'INCLUDE Keyword: "Garnet"'. There is a 'Saved Searches' folder on the right.
- Results Displays:** Contains three sub-sections: 'Top Senders' (listing email addresses like chris.moore@jkenterprises.com), 'Top Recipients' (listing email addresses like chris.moore@jkenterprises.com), and 'Top Phrases' (listing words like transfer, code freeze, compile, Regards,). Below these is a 'Timeline' chart showing activity from 2000 to 2002.
- Results List:** A table with columns for Date, Flags, Sender, Recipient, Subject, and Content. It shows 2,388 results. The first few rows include dates like 2/10/00 and various email addresses and subjects.

Four yellow callout boxes point to specific features:

- 'Search & browse' points to the search filters on the left.
- 'Workflow integration' points to the 'Review Flags' section.
- 'Interactive visualizations' points to the 'Timeline' chart.
- 'Risk analysis from content' points to the 'Top Phrases' section.

A status bar at the bottom says 'User searched for keyword "Garnet"'. A yellow bar at the very bottom of the slide contains the text 'User searched for keyword "Garnet"'. There is also a small yellow bar at the top left of the slide.



# Analysed content provides deep visibility and value in a wide range of use cases

Industry	Content analysed	Enabling visibility into...	Delivering Business Value
Manufacturing	Technician notes, warranty claims	Emerging quality issues	Reduced warranty costs & improved customer satisfaction
Insurance	Claims materials	Emerging claim trends & fraud potentials	Proactive business response to fraud & shifts in claim workload
Finance	E-mails	Compliance violations and legal liability (eDiscovery)	Reduced legal and financial exposure
Law Enforcement	Police records, 911 calls, statements, ...	Rapid crime solving and crime trend analysis	Safer communities and optimized force deployment
Media, FSS, Government	Public opinion sources (news, blogs, forums)	Market perception and sentiment	Improved products & increased marketing effectiveness
Pharma, Life Sciences	Patent and scientific literature repositories	Improved drug development strategy	Higher return on drug R&D investment
Healthcare	Provider, patient, treatment information	Dashboard view of providers	Stronger negotiating position with providers
Telecom, FSS	Customer calls and e-mails	360° view of customer; churn prediction; marketing opportunities	Improved customer retention, agile market responsiveness with new offerings

- **New Content Analytics products and capabilities coming in 2008-2009**

- IBM Content Analyzer available in limited geos; more solutions and WW offerings coming
- Get started today by organising your content with IBM Classification Module



# Agenda

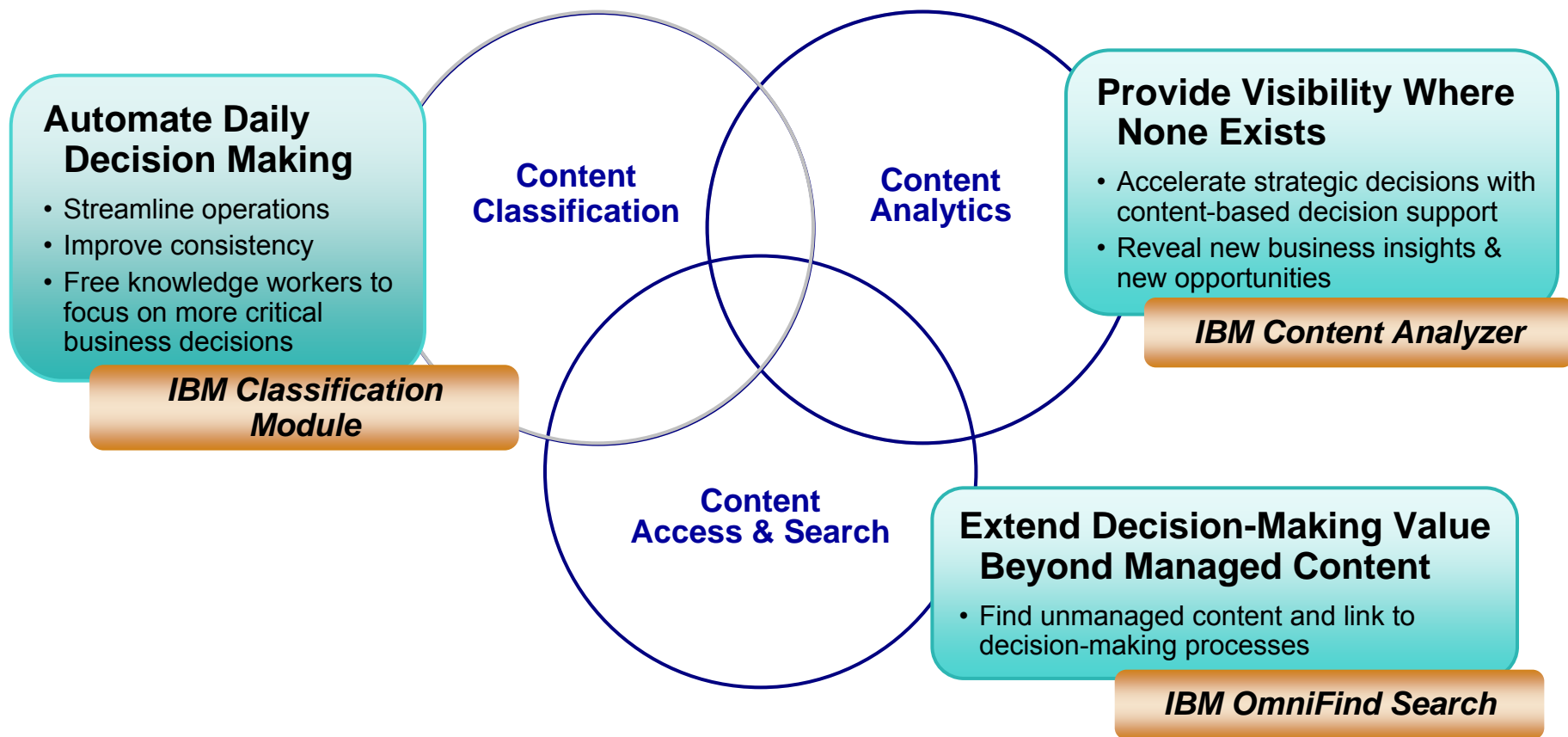
- Introduction
- Automating Daily Decisions: IBM Classification Module
- Extending Beyond Managed Content: OmniFind search
- Providing Visibility Where None Exists: Content Analytics
- ➔ ■ **Wrap-up**



# Recap

- Content Discovery takes ECM to the next level, leveraging your content to drive business optimisation

## ECM Content Discovery Capabilities



# Take Action for Business Optimisation!

- **Get educated:**

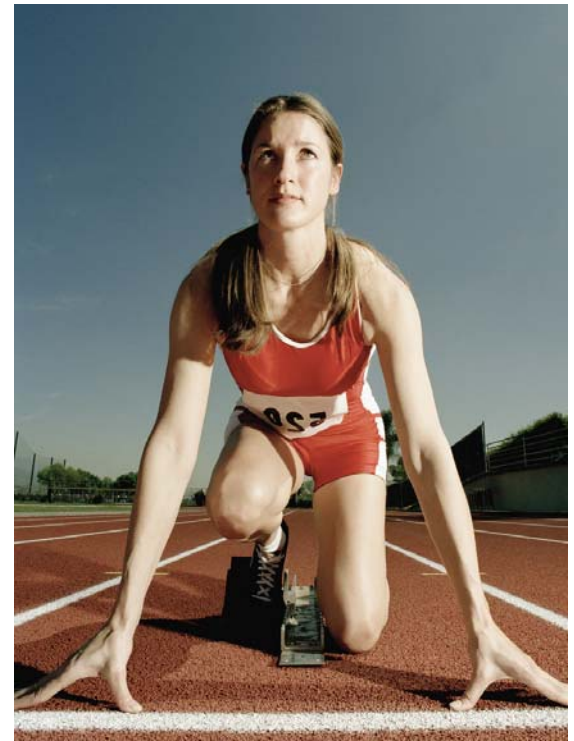
- Classification: <http://www.ibm.com/software/data/enterprise-search/classification/>
- Content Analytics: <http://www.ibm.com/software/data/enterprise-search/omnifind-analytics/>
- Search: <http://www.ibm.com/software/data/enterprise-search/omnifind-enterprise/>

- **Deploy IBM Classification Module to start automating daily decisions**

- **Start planning for the business insight of content-centric BI with IBM Content Analyzer**

- **Leverage OmniFind search to expand insight beyond managed content**

- **Contact your IBM Representative for further information**



Thank  
YOU

**Brendan Reid**  
Global ECM Solutions Specialist  
IBM Information Management Software

