



IBM Almaden Research – Intelligent Information Systems

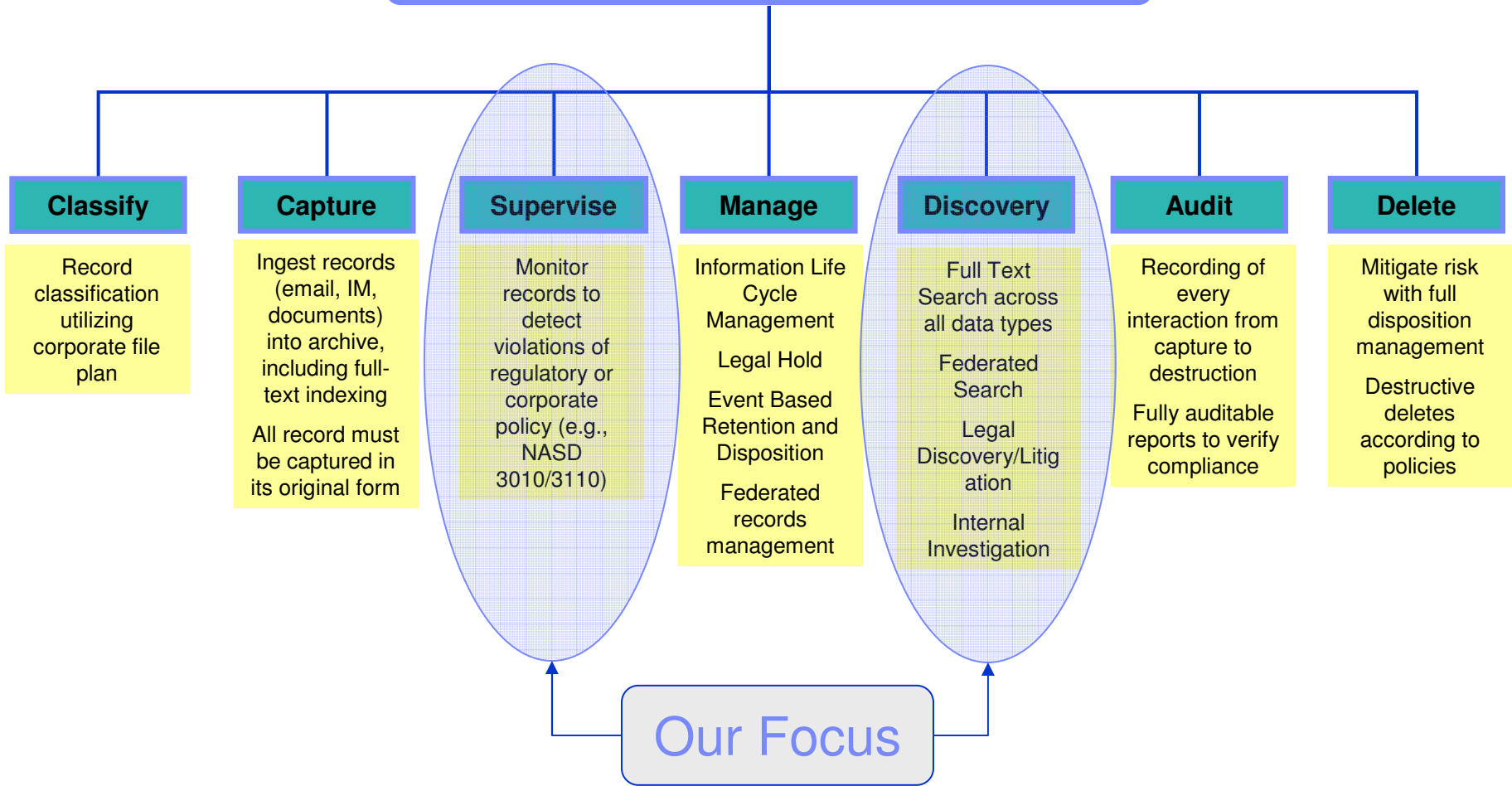
Galaxy: Supervision and Deep discovery for email archives

Stefan Edlund, Joshua W Hui



@business on demand software

Email Compliance Solution



Supervision and Discovery

■ *Supervision*

- ▶ Monitor any email record, before it is received
- ▶ Detect any suspicious activity through pattern matching described by rules, and hold it for further review
- ▶ Hook to the record transfer agent (e.g. email server)
- ▶ Hook to the record ingestion processing

■ *Discovery*

- ▶ Provide search capability to the archive records, mainly for
 - Litigation/Legal search
 - Detect any suspicious pattern
 - Internal investigation



Regular Discovery VS. Compliance Discovery

- Regular Discovery
 - ▶ Find a piece of information, such as
 - *Find a particular product fitting my search criteria?*
 - *Find a review for a given restaurant?*
 - ▶ As long as **ONE** result answers the question
- Compliance Discovery
 - ▶ Need to understand all the data.
 - ▶ Each piece of information have an **EQUAL** importance
 - ▶ They can relate or link to one another to provide a bigger picture



Discovery component

■ *Discovery*

- ▶ Known limitation to existing products:
 - ➡ Plain search result is *only enough* for litigation, but *not* for deep discovery
- ▶ Need better ways to organize or understand the data
 - Different perspectives on the search result,
 - e.g. email distribution over time, Top-N senders/receivers
 - Multi-facet search interface for drilling into different dimensions
 - Time event analysis
 - Semantic search
 - use text-mining (classification) and UIMA-based annotation



Galaxy Discovery Interface

Drill-down Capability

Social Network

The screenshot displays the Galaxy Supervision and Discovery Client interface. At the top, there are navigation tabs: Search, My Work Items, Rules, Log, Network, Term Analyzer, and Temporal Analyzer. Below these is a search bar containing 'easy AND sell' and a 'Run' button. The main area is divided into two panes. The left pane shows a table of search results for the query 'class:"business"', displaying 33 of 1014 results. The right pane shows a social network graph with nodes representing email addresses and edges representing relationships. A 'Summarization' label is on the right side of the graph.

#	From	Subject	Date	Score
1	m..schmidt@e...	Enron Mentions	2001-10-26:06.16	0.00000
2	m..schmidt@e...	Enron Mentions	2001-10-26:14.45	0.00000
3	m..schmidt@e...	Enron Mentions	2001-10-25:06.17	0.00000
4	domesticatio...	Party Time	2002-02-26:14.45	0.00000
5	e.bulletin@e...	Maximize your 401...	2002-01-16:18.04	10.00000
6	james.steffe...	Enron Proposed Re...	2001-04-09:12.42	14.05480
7	ann.schmidt@...	Enron Mentions	2001-04-10:00.48	0.00000
8	ann.schmidt@...	Enron Mentions - ...	2001-04-16:00.50	0.00000
9	ann.schmidt@...	Enron Mentions	2001-04-19:00.50	0.00000
10	ann.schmidt@...	Enron Mentions	2001-04-16:09.48	0.00000
11	owner-evenin...	Sea Power & Assoc...	2001-04-16:10.53	5.56478
12	ann.schmidt@...	Enron Mentions	2001-04-17:00.41	0.00000
13	ann.schmidt@...	Enron Mentions	2001-04-17:09.59	0.00000
14	ann.schmidt@...	Enron Mentions	2001-04-11:09.44	0.00000
15	e.bulletin@e...	April e.Bulletin	2001-04-11:11.26	0.00000
16	james.steffe...	Enron Proposed Re...	2001-04-12:13.58	14.05480
17	yahoo-delive...	Yahoo! Newsletter...	2001-11-21:17.36	0.00000
18	jeff.dasovic...	Comments on Denni...	2001-03-21:02.43	0.00000
19	awu@quinnema...	Re: Comments on D...	2001-03-21:04.41	0.00000
20	jeff.dasovic...	Comments on Denni...	2001-03-21:07.22	0.00000
21	scott.stones...	Re:	2001-04-04:09.08	5.56478
22	ann.schmidt@...	Enron Mentions	2001-03-22:23.41	0.00000
23	ann.schmidt@...	Enron Mentions	2001-03-28:00.35	0.00000
24	ann.schmidt@...	Enron Mentions - ...	2001-04-02:01.29	0.00000
25	ann.schmidt@...	Enron Mentions - ...	2001-04-02:02.35	0.00000

Search Result

Summarization

THE END

- <http://www.almaden.ibm.com/cs/projects/iis/galaxy/>

