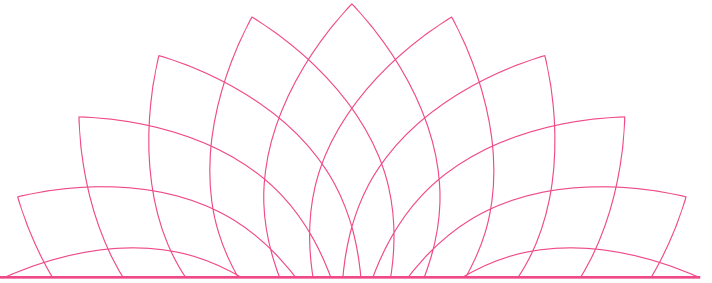




# SINTELIX



## CASE STUDY

# National Archives Enterprise Information Management

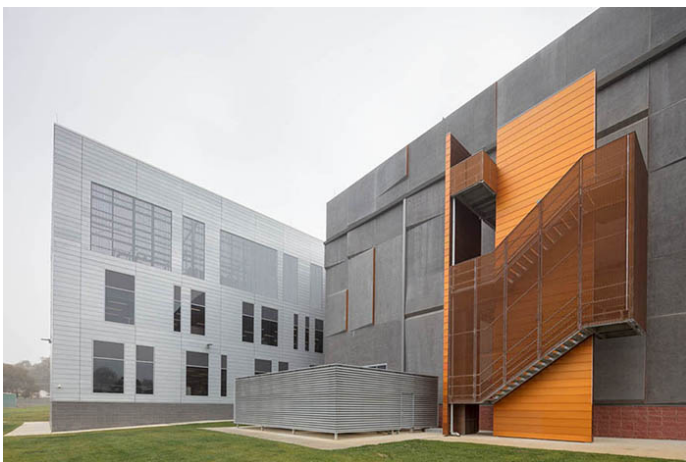
**“We pay tribute to the company’s ability to analyze a difficult business problem and come up with a viable and effective solution.”** Tatiana Antsupova, National Archives of Australia.

## The Client: National Archives of Australia



The National Archives of Australia (NAA) can best be described as the memory of our nation – collecting and preserving Australian Government records that reflect our history and identity.

Their collection traces events and decisions that have shaped the nation and the lives of Australians. As well as preserving our history, the National Archives plays a key role in helping to ensure the Australian Government and its departments are effective and accountable to the people.



National Archives of Australia Preservation Facility

## The Challenge:

### Automated extraction of metadata and classification of records

NAA staff were manually appraising and disposing of records, a very time consuming, expensive and inconsistent process. NAA required a robust system for automatically recommending disposal actions using document content and metadata. The image below shows the complexity of the manual process.

Semantic Sciences responded with an ambitious proposal to develop a process and technology that aimed to close the wide gap between general disposal authorities such as AFDA Express (a streamlined version of the Administrative Functions Disposal Authority AFDA) and the automatic document sentencing. The proposed enterprise information management tool would support automatic extraction of relevant metadata and auto-classification of records for document sentencing and disposal.

The core challenges Semantic Sciences addressed were:

1. The need to create **metadata** for documents by automatic inference;
2. Creating **auto-classifiers** that require little training data, are accurate and whose operation is understandable by people;
3. The need for a simple methodology and technology for creating **“gold standard”** sentencing decisions for sets of documents which could be used to train and assess the performance of the structured auto-classifier.



## The Solution: Enterprise Information Management Tool

Semantic Sciences provided a solution that split the overall complex problem into several straightforward activities, with simple interactive tools for each one. The approach was fast, testable and accurate.

Developed by Semantic Sciences, the Sintelix text and data analytics platform delivered a lot of the capability required by NAA. To support iterative development of automatic sentencing, Semantic Sciences supplemented Sintelix's existing capabilities with three innovations:

1. A specialised taxonomy viewer/editor for layered taxonomies.
2. A highly productive search-based process for rapidly creating "gold standard" allocations of documents to nodes on the taxonomy tree; and
3. A development system for "structured classification" which coherently combines a network of simple classifiers to determine document sentencing.

The screenshot displays the Sintelix auto-classifier tool interface. On the left, the 'Class (Classification: Auto-classifier)' configuration panel is visible, showing three 'Rule Generators' with their respective data sources and cosine metric weighting options. The 'Learned Classification (Early Stage)' section is active, indicating that classification is fully handled by the learned classification process. On the right, the 'Summary of Inputs' and 'Evaluation: Disposal...' and 'Evaluation: Class' tabs are shown. The 'Training Partition' and 'Testing Partition' tabs are selected, displaying an accuracy of 83.3% and 48 documents evaluated. A 'Score Breakdown' table is present, showing performance metrics for various tags. Below the table, the 'Classification Stages' section shows the performance of individual stages, including a learned classifier with 13 possible tags, 48 hits, 40 correct, and 8 spurious (100% of all mistakes).

Expected Tag	F1	Precision	Recall	Correct	Spurious	Miss
Overall	83.33	83.33	83.33	40	8	8
Biographies	66.67	100.00	50.00	1	0	1
CEO Ken Lay from to email	100.00	100.00	100.00	12	0	0
Enron Transformation Plans	66.67	66.67	66.67	2	1	1
HR - Personal	50.00	100.00	33.33	1	0	2
Industry Newsletters	94.74	100.00	90.00	9	0	1
Internal Newsletters, Circulars	80.00	66.67	100.00	4	2	0
Invitations, Proposals & Correspondence	80.00	100.00	66.67	2	0	1
Management	0.00	0.00	0.00	0	3	2
Other Annual Reports	100.00	100.00	100.00	1	0	0
Policy & Methods	100.00	100.00	100.00	1	0	0
Projects and Operations	100.00	100.00	100.00	1	0	0
Research, Teaching	85.71	75.00	100.00	3	1	0
Risk Management Products	85.71	75.00	100.00	3	1	0

National Archives of Australia auto-classifier tool by Sintelix



**The Outcomes and Benefits:**

The final solution delivered two key outcomes:

1. **Deployment of the Enterprise Information Manager (EIM)** software which provides a user interface to existing Sintelix capability to enable productive and accurate BCS design, automatic record sentencing and integration with their electronic document records management system (EDRMS).

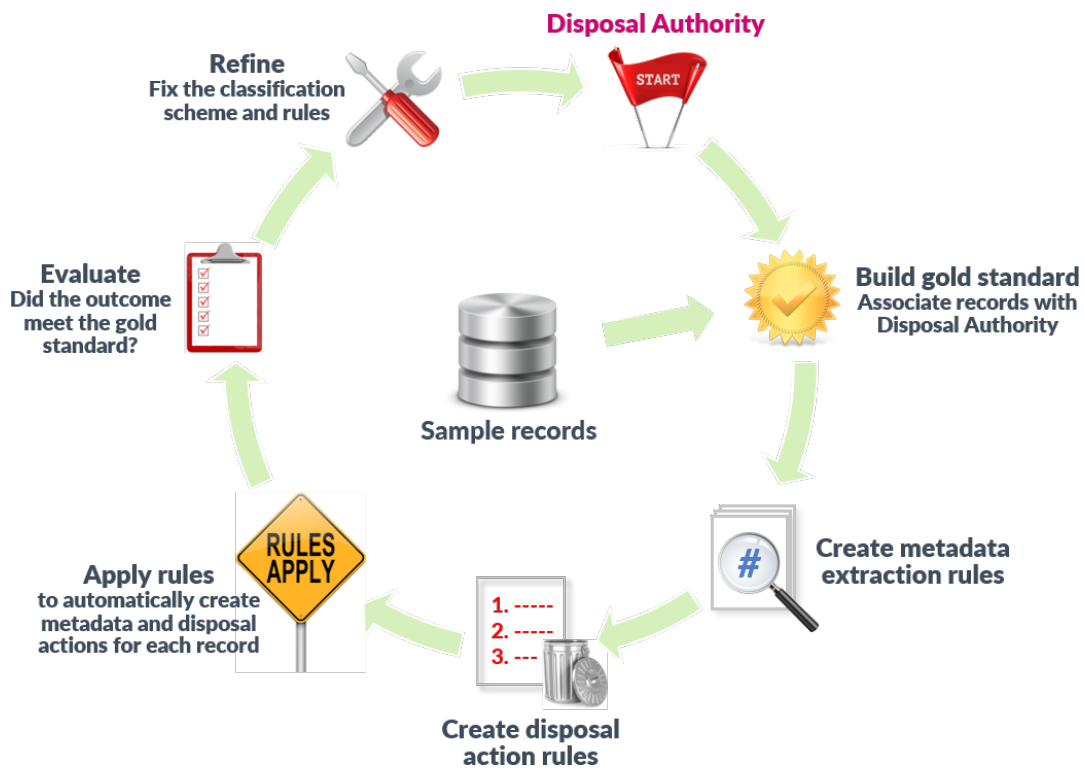
2. A tool to assist NAA with the development of **automatic classifiers** (including business rules) for sentencing records and information.

The image below shows how Semantic Sciences developed a "gold standard" solution for document sentencing and disposal.

The resulting system (Enterprise Information Manager) is able to store and sentence large volumes of documents at high speed. It allows Australian Government Agencies to implement automatic document sentencing tailored to their needs. The system permits human intervention but provides consistent results. It is orders of magnitude faster than manual approaches.

*"The project was completed in two months. We found working with Semantic Sciences to be interesting and rewarding. We pay tribute to the company's ability to analyze a difficult business problem and come up with a viable and effective solution. I would welcome the opportunity to partner with Semantic Sciences again on future projects"*

Tatiana Antsupova, Assistant Director, Digital Strategy and Solutions, National Archives of Australia.



Gold standard solution for document sentencing and disposal

**Discover Sintelix**



**Video**  
[sintelix.com/demo-video](http://sintelix.com/demo-video)



**Demonstration**  
[sintelix.com/demo-request](http://sintelix.com/demo-request)