



Australian Government
Department of Finance



Digital Records Transformation Program

Industry Market Day

Findings

26-29 September 2017

Contents

Purpose	3
Background	3
Scope	3
Approach	3
Key Findings	4
Artificial Intelligence (AI)	4
AI Maturity Analysis	5
Open Source Technologies	5
Open Source Technologies Maturity Analysis	5
Microservices Architecture (MSA)	6
MSA Maturity Analysis	6
RESTful API	6
RESTful API Maturity Analysis	7
Ontology Management	7
Ontology Management Maturity Analysis	7
Cloud Maturity	8
Cloud Maturity Analysis	8
Records Management	8
Records Management Maturity Analysis	8
Appendix 1 – Proposed Platform approach	10

Purpose

In late September 2017, interested parties were invited to present at a series of Market Days on their product offerings and alignment with the proposed approach for a whole of government platform as outlined in the *Position Paper on the Indicative Strategy for the Development of a Whole of Government Digital Records Platform* (the Position Paper).

The purpose of this report is to summarise the findings of the Market Days exercise. These findings have assisted the Department of Finance (Finance) in planning its delivery approach for a records management platform as part of the Digital Records Transformation Program.

This report presents broad findings against the proposed approach for a whole of government records solution as outlined in the Position Paper. The information included is based on the Market Day presentations and accompanying notes provided by interested parties. It is not intended to be a comprehensive summary of all Market Day presentations.

Background

In early 2016, Finance undertook a Request for Information activity, with the purpose of understanding innovation within the records management industry. The results of this activity indicated that industry was not yet developing a service capable of automating records management in the Australian Government context.

In 2017, the Digital Records Transformation Program received funding as part of the Australian Public Service (APS) Modernisation Fund. Government agreed that the aim of the Program is to modernise the common function of recordkeeping across the APS by taking advantage of new technologies, particularly those that automate the capture and classification of records.

Scope

In late 2017, Finance hosted a series of Market Days to investigate the records management industry's capabilities to provide a whole of government platform based on the proposed approach found in the Position Paper (refer Appendix 1).

Interested parties were invited to submit a five-page document outlining their capabilities. These submissions were reviewed to inform the facilitation of the Market Days. The analysis of the submissions and presentations did not form part of a procurement exercise, and product and vendor capabilities were not assessed in anticipation of a tender.

Approach

Finance released the Position Paper to the records management industry in August 2017 via AusTender. Following the release of the Position Paper, Finance held an open industry briefing where the industry was encouraged to ask questions of the Digital Records Transformation Program. Attendees, both online and in person, exceeded 150.

The industry briefing was followed by the industry Market Days. Finance invited all interested parties to present on their capabilities, followed by a question and answer session. Twenty-eight vendors registered for the Market Days, which were held from 26 – 29 September 2017 at the Department of Finance.

Each Market Day presentation was to a panel comprised of program members and cross-agency subject matter experts.

Key Findings

Key strengths demonstrated by interested parties during the Market Days included: connectivity to external systems; automation through rules-based approaches; and, a move towards microservices architecture. The industry is also beginning to develop some support for machine-learned auto-classification capabilities, which are key to the proposed approach for the Digital Records Platform (DRP).

There was little evidence of the application of artificial intelligence (AI) or ontology management capabilities.

As a result, the DRP will focus on testing the feasibility and demonstrating the benefits of employing automation and semantic technologies in a Demonstration of Concept.

The Market Day findings are consistent with those of Mariano and Woodbridge (2017) of Gartner, who surveyed 14 of the most popular content services and records management vendors, identifying whether and how they are expanding their capabilities through automation and use of evolving and disruptive technologies such as AI technologies¹. The Gartner survey found that most vendors' automation capabilities are through rule-based classification and not AI technologies. The Gartner survey also stated that a high number of the vendors are investing in AI technologies for records management, which indicates that while vendors are moving toward use of AI, they are not yet at production stage with their capabilities.

Artificial Intelligence (AI)

AI, and more specifically, auto classification is a foundational requirement of the DRP's innovation. The DRP should be able to create customised capabilities for auto classification of records using state-of-the-art machine learning algorithms.

Once operational, the DRP will need to be able to ingest large amounts of data. To minimise reliance on human intervention for routine records management tasks, basic tasks will need to be automated through machine learning algorithms and refined through user analysis techniques.

¹ Mariano, J. & Woodbridge, M. (2017) *Evolving and Disruptive Technologies Are Critical to Streamlining Records Management Programs*, Gartner, <https://www.gartner.com/doc/3717917/evolving-disruptive-technologies-critical-streamlining>

The volume of data to be stored in the Platform will grow significantly every year. By using technologies more commonly associated with linked data and data analytics, it is possible to ingest and manage large volumes of data in the solution. The scaling of automation for records classification can also be achieved through incorporating these technologies. Without trialling these technologies, running a machine learning classifier that can auto-classify records of the expected volume could take days and even months.

AI Maturity Analysis

AI maturity was considered against an interested party's ability to demonstrate AI or the beginning of AI adoption, through: research and development; rules-based algorithms; supervised and unsupervised machine learning; and, complexity and applicability of the learning algorithms.

While many of the products presented at the Market Day demonstrated capabilities outlined in the Proposed Approach, around a third did not reveal maturity in AI capability.

Several interested parties indicated that while they do not currently have AI capability in their products, they are looking to this in the future.

Other interested parties' products demonstrated the use of rule-based algorithms. Of these, several were also clearly applying early AI for classification purposes. Encouragingly, a small number of interested parties had included supervised learning with more complex algorithms for document management processes.

Open Source Technologies

The DRP is seeking to use Open Source Technologies as much as possible. This aligns with the *Digital Service Standard* and other government initiatives including the National Archives of Australia's *Digital Continuity 2020 Policy*.

Open Source Technologies are widely used by scalable solutions and have a strong developer community behind them.

Open Source Technologies Maturity Analysis

The approach outlined in the Position Paper sought interested parties' use of Open Source Technologies, acknowledging that fully Open Source products are not common. Open Source maturity was considered against: the ability of products to produce outputs that can meet Open Source standards; documentation of tools and technologies incorporated; and, the proportion of the products' components that are open source and proprietary.

A large proportion of interested parties demonstrated an initial maturity, noting that in the future they may incorporate Open Source Technologies into their products. A small number of interested parties did not demonstrate Open Source maturity.

Almost a third of products demonstrated by interested parties could produce documents in Open Source formats; others were able to integrate some standardised Open Source Technologies. A small number of interested parties demonstrated that the majority of their products were Open Source, with minimal proprietary components.

Microservices Architecture (MSA)

The DRP is intended to incorporate best practice MSA, enabling the platform to demonstrate high scale interactions with external systems. The DRP's interactions with external systems would include:

- systems acting as API consumers – these are high volume, with low data exchange interactions; and
- external data storage – these are low volume, but high data exchange interactions.

To ingest records and extract metadata, the DRP is intended to interact with external systems (API consumer and data storage) via endpoints provided by RESTful APIs.

MSA Maturity Analysis

Interested parties' products were considered against a number of principles and maturity based on aggregation of the principles. Principles for MSA maturity are:

- Stateless Microservices where all state related information is stored in a cache cluster;
- Central logging, where all information about Microservices, their interactions and health status is stored in a central logging cluster;
- Componentization via Services – each service is equivalent to a component that is the smallest replaceable and upgradable building block in the solution;
- Smart endpoints and dumb pipes – services are as decoupled as possible;
- Horizontal Scaling and automation, where the architecture supports horizontal scaling in and out and failure of an instance of a microservice does not affect the solution, also services can be tested and deployed automatically.

The Market Day indicated that MSA maturity in the market is mostly at an initial stage. Many interested parties had greater maturity in RESTful APIs and indicated that they were moving to offering microservices architecture.

RESTful API

The DRP must support and enable developers to add required APIs. This will allow the DRP to be extensible and will future proof the DRP from integration and scalability perspectives. Support for MSA heavily relies on RESTful APIs, hence it is necessary to understand the maturity of RESTful APIs in the market, including interactions and how much of the APIs have been exposed externally for future integrations.

Often in a proprietary solution, the source code is closed and information for records management may be inaccessible unless being accessed through the proprietary solution. This means that documentation to work with the solution is not available and there is a chance of vendor lock-in. The RESTful API architecture supports extensions and open standards, which assists in avoiding such a scenario.

RESTful API Maturity Analysis

Largely, the Market demonstrated high maturity for RESTful API capability. Over half of interested parties could demonstrate a high level of maturity. Almost a third of interested parties had developing RESTful API maturity with some deployment and limited documentation. Several interested parties demonstrated well-defined maturity in that their products support RESTful APIs with full documentation available. This bodes well for the DRP, which is seeking to have a full microservices architecture built with RESTful APIs.

Ontology Management

Using semantic technologies and ontology management permits a formal and explicit view of government entities and their activities. This means that it is possible to manage changes that occur in policy and structure (e.g. machinery of government changes), as well as societal changes. These changes are represented in human and machine-readable ways, thus assisting in making better use of government information.

The Program has been developing strength within the team for ontology management and design for government records management (the Australian Government Records Interoperability Framework (AGRIF) Project).

The AGRIF is designed to be a system of related ontologies that permit Information Managers in different government entities to describe their own local structure and activities while remaining conceptually interoperable and discoverable in the broader context of the Australian Government.

For the solution to be able to support Ontology Management and enable use of the AGRIF, graph technology is necessary. Records in the DRP will need to be extensively interconnected and have a graph data structure. As a result, they cannot be represented effectively by using traditional relational databases or the NoSQL data stores. To elaborate, technologies with graph support address these requirements because they support:

- horizontal scalability;
- graph style interconnected relationships and are less cognitively taxing than other representations; and
- sophisticated queries for data extraction with sub-second response time.

Ontology Management Maturity Analysis

Interested parties' products were considered against a number of principles and maturity based on aggregation of the principles. Principles for Ontology Management ² maturity are:

- understand and use Web Ontology Language (OWL2);
- store ontology as a graph, in a graph database;
- import, export, ontologies via web;
- allow for automatic validation (e.g. of logic, by reasoners); and,

² These Principles are loosely based on World Wide Web Consortium recommendations.

- support management by information management users via an interface.

A number of interested parties demonstrated consistent maturity, meaning they satisfied at least three of the principles for ontology management.

The majority of interested parties were in the initial maturity stage. Almost a quarter were not applicable, as their products were never intended to have an ontology management capability.

Cloud Maturity

While Platform as a Service Providers (PaaS) and Software as a Service (SaaS) Providers offer several services that can be used to build the DRP's PaaS and SaaS capabilities, products that heavily rely on them may deliver solutions that will have difficulty migrating to different infrastructure.

Initially, DRP will need to support both unclassified and classified information, while acknowledging only a limited number of cloud providers can support an environment for Protected Level information.

The DRP should be delivered with PaaS capabilities that do not rely on features and services provided by any specific cloud provider. This will greatly simplify the migration process to an infrastructure that can support Protected Level government information, when available.

Cloud Maturity Analysis

The Market demonstrated it is rapidly moving to support delivery of services through cloud technologies that have minimum dependence on single providers. A small number of interested parties demonstrated the capability to integrate systems with products and technologies that not only support horizontal scaling, but also auto deployment and private-cloud installation abilities.

The majority of interested parties were developing maturity and had a medium dependency on PaaS providers. Almost a third of interested parties demonstrated heavy reliance on a single PaaS provider. A small number of interested parties' products did not include any cloud capability.

Records Management

The Australian Government has a records management environment regulated primarily by the National Archives of Australia. Any system to be used for records management within the Australian Government should be able to meet, at minimum, the requirements of ISO 16175 – *Principles and Functional Requirements for Records in Electronic Office Environments*, and ISO 15489 – *Records Management*.

Records Management Maturity Analysis

All interested parties had some level of records management maturity for the Australian Government environment. Several interested parties not only understood requirements and

standards for the Australian Government, but also were also able to demonstrate them to a suitable level.

The majority of the market demonstrated, at minimum, an awareness of the records management requirements and understanding of the standards and guidelines that support the requirements.

DRAFT

Appendix 1 – Proposed Platform approach

- a. use RESTful microservices for information management (lifecycle, search, sharing, sentencing, disposal) in a service-oriented architecture configuration;
- b. use World Wide Web Consortium (W3C) recommendation-compliant linked-data approaches for description and discovery;
- c. employ context and content extraction techniques to enhance description, discovery, management, and archival activities with automation;
- d. use cloud-based approaches, with data maintained onshore, in order to present a SaaS-like offering to Australian Government agencies;
- e. use open source technology stack where possible;
- f. potentially implement a cloud hosted Application Programming Interface (API) gateway service;
- g. integrate the microservices with the Department of Industry, Innovation and Science's VANguard federated identity system for client authentication, via OAuth2;
- h. establish an automated deployment framework that is Infrastructure-as-a-Service platform independent (e.g. using Docker or a similar framework);
- i. develop three application integration plugin reference implementations (for Windows File System, Outlook, and SharePoint);
- j. develop reference implementations for keyword extraction and automated sentencing algorithms;
- k. iteratively tune and refine the keyword extraction and machine learning algorithms using pilot agency sample data sets;
- l. employ the Australian Government Records Interoperability Framework (the framework uses OWL2 – see Attachment D);
- m. align with the Digital Service Standard;
- n. ensure that the solution is capable of processing all records in its first year from pilot agencies, scaling to all government records in unstructured systems in Year five;
- o. ensure that the solution is able to meet each user agency's business continuity requirements;
- p. meet the Australian Government's Information Security Manual requirements and pass an Information Security Registered Assessors Program (IRAP) assessment; and
- q. in the future, support third-party markets for providers of connector services, algorithm-based automation services, and user interfaces.