



---

# FAA E-Discovery

## A Cloud Computing Use Case

Version 0.4 - DRAFT

## Document Revision History

Revision	Date	Change	Author	Company
0.1	5/19/2011	Initial Creation	Viktor Kaufmann	Knowcean Consulting
0.2	7/14/2011	Added comments provided by members of the FAA team.	Viktor Kaufmann	Knowcean Consulting
0.3	9/14/11	Added additional feedback and comments.	Viktor Kaufmann	Knowcean Consulting
0.4	10/21/11	Additional edits.	Paul Fontaine Viktor Kaufmann	FAA Knowcean Consulting

## Disclaimer

This document has been prepared by the National Institute of Standards and Technology (NIST) and describes standards research in support of the NIST Cloud Computing Program. Certain commercial entities, equipment, or material may be identified in this document in order to describe a concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that these entities, materials, or equipment are necessarily the best available for the purpose.

## Table of Contents

Disclaimer .....	2
Description .....	4
Background.....	4
Cloud Computing Concept of Operations .....	6
Analysis.....	7
Service Model.....	7
Deployment Model.....	7
Cloud Computing Essential Characteristics .....	8
On-demand self-service.....	8
Broad network access.....	8
Resource pooling.....	8
Rapid elasticity.....	8
Measured service .....	8
Security.....	8
Interoperability.....	9
Portability .....	10
Maintainability .....	10
Usability.....	10
Performance.....	10
Resilience.....	11
Concerns and Challenges.....	11

## Description

The Federal Aviation Administration (FAA) is examining how to implement cloud-based e-discovery and Freedom of Information Act (FOIA) processes for email. The system must be able to perform discovery in both its in-house email implementation (Lotus Notes) but also in cloud-based email systems. The system will also be used to manage content for compliance purposes, and will serve as an archive of FAA messaging content.

The long-run goal is to support four primary functions: e-discovery, electronic records management, FOIA, and privacy. These four processes have similar needs and capabilities, including searching business applications, document repositories, email (including calendar, contacts, tasks, etc.) and instant messages, and distributed storage (both internal and external) for electronically stored information (ESI) meeting defined criteria. The focus of this business use case is the processes and systems required to respond to e-discovery and FOIA requests as they pertain to email message data and other supporting data such as calendar entries, tasks, attachments, etc. that are produced and processed by the FAA's traditional and cloud email messaging systems.

E-discovery software must retain and report metadata within the archives. For example, email content metadata such as the folder an email resides in, any attachments to the email, whether the email is a reply to other emails or has been replied to, must be retained. Notes, contacts, tasks, calendar entries, and attachments need to be recognized and captured. Searches of both cloud and hosted email systems must be able to access both the object and any attachments to that object (for example, a search would need to be able to identify a relevant document attached to a calendar entry).

## Background

E-discovery requires identification, preservation, collection, processing, review, and production of ESI<sup>1</sup>. ESI includes not simply email and attachments, but also files, instant messages, web pages, databases, and other data stored on computers, networks, backups, and other storage media. Metadata such as attributes, movements, and storage locations are collected as well. As stated previously, this use case considers the specific case of messaging content and its associated metadata.

ESI from readily accessible locations such as hard drives and active network servers, removable media, and sources sent for storage (as opposed to backups) typically satisfy e-discovery requests. It is usually not required to search the backup systems or find deleted files. For email,

---

<sup>1</sup> A more detailed definition of electronic discovery can be found at <http://www.edrm.net/resources/glossary/e/electronic-discovery>.

this would include the central email locations and any archives created by users on local machines.

E-discovery provides for discovered files to be produced in various formats that are negotiated between the litigating parties. These formats can include the native, TIFF, and PDF formats.

The e-discovery process goes through several steps:

- Notification to or awareness by the organization that potentially relevant information must be retained.
- Implementation of a litigation hold, which informs relevant custodians and administrators of their duty to preserve potentially relevant information.
- Identification of ESI that meet required search criteria. These files can reside on file systems, workstations and laptops, and peripherals. These candidate files are then copied to a secure location on the network to prevent modification.
- Review of ESI for relevancy or privilege.
- Production of ESI that were determined to be responsive to a document request.

The search functionality needs to be able to identify files that are in readily accessible locations.

- The tool must be able to read encrypted documents using encryption software provided by the FAA. The tool would be provided a master key to enable this functionality. Ideally, individuals encrypting email will be prevented from using their own encryption software; instead, they will be required to use agency-supplied software.
- The tool needs to be able to extract data from the email server or servers and move it to an archive to enable review of candidate files. It is expected that a standard API will be required to allow the tool to successfully search multiple email systems. The FAA voluntary consensus standards<sup>2</sup> for API are preferred over proprietary standards such as MAPI.
- The system must be able to access or import local archives created by users on their local machines, moving candidate files to the result archive.
- The system must be able to search or ingest backup tapes, moving candidate files to the result archive.

The FOIA process is similar to the e-discovery process, with nine exemptions<sup>3</sup>:

- Secret national defense or foreign policy documents;
- Internal personnel rules and practice documents;
- Documents exempted by other laws;
- Confidential commercial or financial information obtained from a person;
- Privileged inter- or intra-agency memorandum or letter;
- Clearly unwarranted invasion of personal privacy;

---

<sup>2</sup> [http://www.whitehouse.gov/omb/circulars\\_a119](http://www.whitehouse.gov/omb/circulars_a119)

<sup>3</sup> [http://www.epa.gov/foia/foia\\_exemptions.htm](http://www.epa.gov/foia/foia_exemptions.htm)

- Documents for law enforcement purposes;
- Reports about financial institutions regulated by the SEC;
- Documents containing exempt information about gas or oil wells.

The FAA currently uses Lotus Notes as its on-premise email system. While the FAA does not currently have an automated e-discovery and messaging solution in place, discussion of the features of an on-premise solution will aid in understanding how a cloud-based solution differs. For purposes of this business use case, all references to ESI should be taken to refer to messaging content and functionality (including attachments) as generally understood to be part of an email system.

Users of an email system that is integrated with an e-discovery solution are broken into two groups with the following capabilities:

- General Users
  - General users perform standard email activities within Lotus Notes such as composing and sending email, replying to email, classifying email, etc.
- Legal Users
  - Search and browse emails and attachments across user mailboxes;
  - search by dates, keywords, custodians, organizational entities, and other search methods;
  - apply and remove legal holds; and,
  - export a collection of emails to a standard format, such as PDF, .NSF, .PST, Concordance load file with accompanying native files, and/or EDRM-XML.

Legal users currently use manual processes to obtain copies of user mailboxes and network storage for manual searches for purposes of e-discovery and FOIA.

While there are several approaches to traditional e-discovery solutions for email, the most desirable solution performs two major functions in addition to its search capability. First, the e-discovery tool monitors incoming and outgoing email at the gateway, and applies policy based rules in moving copies of email to an archive. This requires that the tool be configured with the master key for the email passing through the gateway so that the email could be decrypted. Second, the e-discovery tool monitors the server, capturing changes that occur and retaining the metadata. For example, it would capture folder structure changes, folder moves, email moves, etc.

## Cloud Computing Concept of Operations

The FAA, in conjunction with Federal Cloud Computing Working Group (FCCWG) and other federal agencies, is seeking a cloud-based e-discovery solution, motivated by the move of email to the cloud. This solution would be composed of an archive, identification and collection capability, data preservation capability, and the processing and export of content. The objective is to implement a cloud-based e-discovery solution that can analyze both in-house and cloud-based email systems.

A cloud-based e-discovery system would need to meet the same requirements as the hosted system described in the previous section. As it is envisioned that the FAA will be moving its email to the cloud, the cloud-based system has additional requirements.

- The tool needs to be able to access any cloud email systems to perform its search, and move the candidate files to the review archive.
- The e-discovery system's archives will be based in the cloud, and candidate files will be moved to that cloud-based archive.
- The storage available in the cloud will expand and contract based on current archive requirements.

The security section discusses security in greater depth. It is expected that certain essential security requirements will be addressed:

- All users will access the system through the FAA network. If remote users are permitted, they would connect through VPN using two-factor identification.
- The application will support CACs (common access cards) and two-factor authentication such as SecureID for access when the user is not connected to the FAA network.
- The solution will be integrated with FAA's LDAP implementation for authentication and authorization.

Cloud-based e-discovery systems need to be able to capture metadata from both cloud-based email services and from traditional internal email implementations. Current efforts involve exploring what metadata are available using IMAP.

## **Analysis**

The e-discovery solution meets the criteria of a cloud application. Since many agencies consider email to be a low-barrier implementation, the ability to meet e-discovery and FOIA requirements is an important next step.

## ***Service Model***

The e-discovery system as described is a SaaS application. The cloud consumer is able to create, destroy, and manage individual user accounts autonomously. The cloud consumer does not need to install, patch, or upgrade or backup applications and the associated data, as this is handled by the cloud provider. The cloud consumer is not concerned with the underlying cloud infrastructure or individual application capabilities beyond selecting an application that meets the consumers' needs.

## ***Deployment Model***

The e-discovery system will be deployed in an outsourced community cloud, where the community includes other federal agencies. The objective is to decrease the resources used by FAA's Office of the CIO (OCIO) to support systems unrelated to the core business of the FAA.

The use of a community cloud provides the FAA and sister agencies the ability to more effectively leverage and share resources.

## ***Cloud Computing Essential Characteristics***

### **On-demand self-service**

The FAA can access the service management interfaces, manage user accounts, increase storage capacity, and create new archives for holds, etc., with no assistance from the cloud provider. The FAA can purchase additional accounts in increments as desired. Administrators can easily revoke user access or change user privileges.

### **Broad network access**

The e-discovery cloud application will be accessible using the web. Since attorneys often do their work in remote locations, secure access must be available from outside the FAA firewall.

### **Resource pooling**

The cloud-based virtualized applications can serve multiple consumers, with physical and virtual resources such as storage and processing being assigned and reassigned based on demand. The community deployment model allows more efficient utilization of computer resources. The interest shown by the FCCWG underscores the community aspect of this cloud implementation. Given the potential size of temporary archives, agencies could benefit most from decreased aggregate storage needs.

### **Rapid elasticity**

User accounts can be rapidly and elastically provisioned to quickly scale out and rapidly de-provisioned to quickly scale in. As the demand placed on the environment by customer's changes, computing, storage, and other capabilities can be automatically increased or decreased. From the perspective of the FAA, resources are unlimited and can be purchased at any time. The nature of legal casework requires this elasticity.

### **Measured service**

The e-discovery application would provide metered usage on one or more dimensions, such as the number of total email storage, total storage used by archives, etc. Resource usage is monitored and reported to both the provider and the consumer.

## ***Security***

As with the email system the application will support, the FAA requires that the e-discovery solution be deployed in a FISMA-compliant environment, with a FIPS 199 impact rating of moderate. The cloud-based system would be integrated with the existing LDAP system for authentication and authorization. The system should be able to use either HSPD-12-compliant Common Access Card or a standard username/password for authentication, depending on configuration. Transport Layer Security (TLS) connections and fully functional SSL 3.0 connections need to be supported, and FIPS 140-2 validated encryption is required.

In general, devices used to access the solution will be encrypted (SSL). Additionally, the transport of data between the devices and the system will be encrypted (using TLS or SSL, as above).

The application will enforce data access rules that delineate what data are accessible by whom. Users not in the Legal Group should not have access to search mailboxes or other areas that they have not been authorized to use. Attorneys will only have access to their own case area.

Application audit logs should capture information about executed searches as well as information about the initiator of the search. System audit logs, such as application errors and security breaches, should be provided to FAA at a general level. Detailed information would only be necessary in the event of a security incident. It is incumbent upon the cloud provider to log its audit trail at a level sufficient to perform an incident investigation.

### ***Interoperability***

The e-discovery system must be interoperable with of the FAA's target cloud email service, current traditional email solution, and potentially external E-Discovery services. In the future, if the FAA were to change its email provider, the new provider would need to be supported as well. As discussed previously, mailboxes and email archives located on any number of servers and local machines within FAA network must be able to be scanned by the e-discovery system.

While the FAA has determined that no current open standard supports e-discovery and FOIA applications in existing email systems, it identified three records management standards that might form the foundation of an e-discovery standard. The three standards in question are the Content Management Interoperability Services (CMIS) specification from the Organization for the Advancement of Structured Information Standards (OASIS)<sup>4</sup>, the Records Management Services (RMS) standard from the Object Management Group (OMG)<sup>5</sup>, and the DOD 5015 Design Criteria Standard for Electronic Records Management Software Applications<sup>6</sup>. Overall, the FAA felt that a standard combining and extending CMIS and RMS would be the most desirable path.

CMIS is a vendor-driven standard for content management interoperability services. The FAA determined that this standard's focus was too broad to be immediately useful for use in e-discovery solutions.

RMS is a records management specification, whose definition was led by the National Archives and Records Administration through the Object Management Group. The OMG standard

---

<sup>4</sup> [http://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=cmis](http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=cmis)

<sup>5</sup> <http://www.omg.org/spec/RMS/>

<sup>6</sup> [jtc.fhu.disa.mil/recmgt/p50152s2.doc](http://jtc.fhu.disa.mil/recmgt/p50152s2.doc)

focuses on records management, and not necessarily email; as a result, the FAA feels that the standard falls short of its requirements.

The DOD 5015 is a Department of defense standard focusing on records management. In reviewing the standard, the FAA felt that the standard was cumbersome and a lighter weight version of the standard was desirable.

### ***Portability***

An e-discovery system has not previously been implemented at the FAA, so from the perspective of historical archives created for e-discovery and FOIA, there are no migration concerns. However, as discussed in the Cloud Computing Concept of Operation section, existing email folders and personal archives must be migrated into the e-discovery service, or the e-discovery service needs to be made aware of the existence of these archives.

E-discovery services must retain and report metadata within the archives. The fact that an email is in a specific folder, has an attachment, is a reply to other emails, etc., must be retained. Notes, contacts, tasks, calendar entries, and attachments need to be recognized and captured. Searches of both cloud and hosted email systems must be able to search both the object and any attachments to that object (for example, a search would need to be able to identify a relevant document attached to a calendar entry).

### ***Maintainability***

FAA personnel will have administrative capability including provisioning/de-provisioning of users and role assignment. Since the cloud solution will be kept current, including updates, patches, and maintenance, the amount of administrative support on-site will be decreased. The FAA will not need to maintain systems to test updates and patches, as that will be handled by the cloud service provider. The FAA, and other organizations using cloud-based e-discovery software, would be relieved of the need to install maintenance patches to ensure compatibility across different email systems.

### ***Usability***

It is expected that a single search request will access all email systems automatically based on options selected at the initiation of the search, without the need for case workers to repeat the request for each email system. This requirement is particularly important because during the migration period from the current eMail system to another, users will be distributed across both systems. The migration period could potentially be six to 12 months or more. While this cloud-based system is not replacing a current hosted solution, it is expected to have similar functionality to hosted solutions.

### ***Performance***

Performance is expected to be the same as that for a hosted e-discovery solution.

## ***Resilience***

Defined as the ability to reduce the magnitude and/or duration of disruptive events to critical infrastructure, the effectiveness of a resilient infrastructure or enterprise depends upon its ability to anticipate, absorb, adapt to, and/or rapidly recover from a potentially disruptive event. As such, it is expected that the cloud service provider be able to allow users to continue work during and/or soon after a disaster with minimal impact on productivity. For more mundane problems, restoration of lost or accidentally deleted data would be rapidly achieved.

## **Concerns and Challenges**

The primary challenge and concern the FAA has is the ability to search cloud-based email systems. Currently, many cloud providers provide search tools, but they are not designed to be e-discovery tools. Rather, the tools are aimed at data cleansing, and are restricted to only searching email. The tools are not able to search other aspects of an email system, such as calendars or contacts, and the tools are unable to provide metadata for the results of the search.

While there are standards that support records management, there are no email specific standards that permit e-discovery applications to have an open, standards-based interface to email systems. These required standards should support the search of calendars, notes, tasks, and contacts of the email systems. Furthermore, these interfaces need to return metadata as well, providing contextual information (such as the location of the email, attachments, other addresses on the email, calendar, etc), date and time attachments were created, native format, etc.

The implementation of the cloud services themselves may, in some cases, create impediments to e-discovery. In "Privacy Recommendations for the Use of Cloud Computing by Federal Departments and Agencies," the Federal CIO points out that "...the nature of cloud storage (e.g., widely dispersed servers or databases located domestically or even overseas) may complicate the ability to identify, preserve, and retrieve responsive ESI in a timely fashion, further jeopardizing the agency's ability to meet its legal e-discovery obligations."<sup>7</sup> Even if the location of databases or servers is known at the time of implementation, a cloud provider could move data to other jurisdictions without notifying an agency, which could affect privacy and e-discovery.

---

<sup>7</sup> Page 8.