



Alfresco – A Next-Generation Platform for Customer Communication Archiving and e-Presentation

Migrate your legacy customer communications archives to Alfresco.

Executive Summary

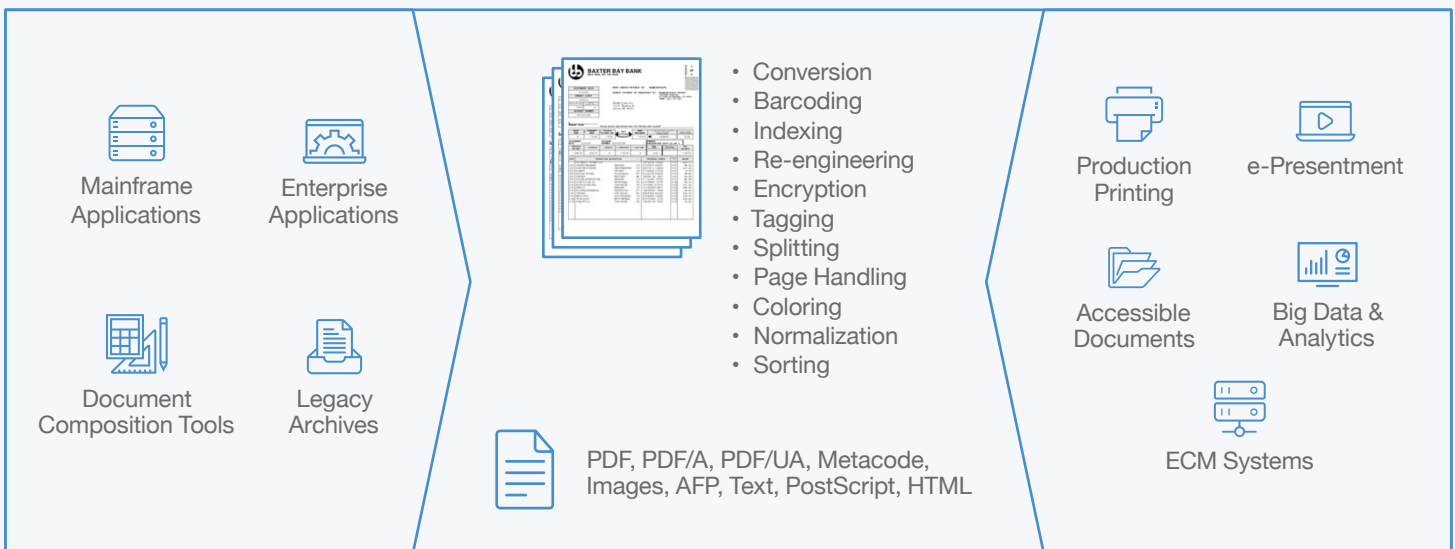
Customer communications are fundamental to organizations of all sizes. They provide a critical means of communicating money owed, account history, case updates, policy details, and many other kinds of marketing and transactional detail to consumers. Without transactional documents like invoices, statements, and acknowledgements, most businesses would simply not operate. And for those organizations in particularly document-intensive industries like financial services, healthcare, or insurance, the critical nature of customer communications underpins regulatory compliance, financial disclosure, and legal discovery. The production and distribution of many of these documents are mandated by regulations to communicate the status of the financial relationship between the company and their customers.




Statements and invoices are by far the heaviest form of outbound communications at 68%.¹

Typically, these documents are produced by enterprise applications, enterprise resource planning (ERP) systems, and document composition tools, which are sent to be printed and mailed to customers, agents, suppliers, and other external entities. Legacy line-of-business and print composition systems that create documents are often expensive and time consuming to change.

Enterprise Customer Communication Workflow



¹Association for Information and Image Management (AIIM) and Crawford Technologies. Trends in Customer Communication Archiving, © 2014



The software that captures electronic versions of these documents and stores them for online access is a customer communications archive, and the process of allowing customers to have electronic access to these documents is commonly called e-presentment. Since printing and mailing documents to customers and other external entities is expensive, it is financially attractive to reduce printing and mailing with e-presentment.

British Gas saw digital e-presentment increase from 0% in 2008 to 21% in 2013 and to over 50% in 2018.²

It is common to find customer communication archiving systems deployed over 20 years ago still in place today. These systems provided great service for many, many years but were monolithic solutions, architected for mainframes and client server architectures and not for the modern, flexible cloud and hybrid-cloud environments organizations want today.

One big challenge for these systems was, and continues to be, the volumes they need to store. Large companies count customer communications in the hundreds of millions, if not billions, which means archives can grow to enormous sizes. The logistical challenge of storing, indexing, and retrieving this kind of content has for a long time been left to these highly specialized IT systems.

Not only are customer communications generated and stored in high volumes but they are usually regulated. For example, in most cases bank statements need to be retained for 7 years, insurance policy documents for 3 years, and invoice and tax documents for up to 11 years. These retention periods once again make storage and retrieval from archives challenging.

In the last 10 years, there have been seismic changes in the way companies communicate with customers through digital channels, and it is no surprise to learn that web, mobile, and email delivery are at the center of this change. Where once customer communications archives served the needs of a small number of users in a call centre, nowadays consumers use self-service tools to access this information directly.

Storing and presenting customer communications may be a necessity, but effectiveness in doing so varies widely. For most organizations, content is maintained by a complex web of systems with content silos from the last 20 years scattered across different departments. The ideal approach has long been defined as a single, centralized platform to manage all information throughout the enterprise, but the central platform has never been particularly successful or realistic, until now.

Fifty percent of organizations have three or more ECM systems, and 22% have five or more.³

As a result, the cost of customer communications services varies widely, and by no means are all companies getting all of the financial benefits of moving to digital communication that they should.

In this paper we describe how Alfresco can be used as the center of a next-generation customer communications archiving and e-presentment strategy and how you can plan for and migrate legacy archives to Alfresco.

²EMC Computer Systems

³AIIM and Crawford Technologies. Trends in Customer Communication Archiving, © 2014

Alfresco – A Next-Generation Platform for Customer Communication Archiving and e-Presentment

So your customer communications archive is old, obsolete, costly, and complex to maintain. It is for those reasons, along with the desire to boost information governance and compliance and to enable better business process management, that organizations decide to make the move to a next-generation document archive.

In this section we look at why Alfresco is an ideal next-generation customer communications archiving platform.

Digital business platform architecture

Alfresco is a re-imagining of the traditional architecture for content management and business process automation, combined into a digital business platform. This means Alfresco includes the concepts and features required to manage content and the business processes that interact in a highly integrated fashion. The content service platform consists of metadata management, full text search, content storage, governance, APIs, and application development tools based on open standards. This enterprise content management (ECM) feature set makes it an ideal target for customer communications archiving.

Scaling the Alfresco platform with AWS

Alfresco has spent a great deal of time and effort to ensure that the Alfresco platform can perform at scale. Benchmarks for over a billion documents have shown that Alfresco is capable of servicing mission-critical customer communications archiving and e-presentment workflows. Optimised for Amazon Web Services (AWS), Alfresco scales even better in the cloud with effectively unlimited storage and indexing capabilities.

Customers using Alfresco customer communications archive and e-presentment systems can now benefit from



the cost efficiencies and scaling of Amazon cloud services instead of tying up expensive mainframe and system resources in on-premises systems.

Document types and metadata

Customer communication types can be varied and complex and it is not unusual to find archives containing hundreds if not thousands of different record types. These records may contain unique document layouts with different indexing criteria, which need to be replicated in Alfresco. Alfresco's document-typing model is flexible enough to cope with these requirements and any migration usually presents an opportunity to rationalise content into a simpler domain model.

And of course the advantage of an integrated ECM platform is that all content, not just customer communications, is searchable within the same platform, allowing knowledge workers to be more efficient.

For 40% of organizations, search across multiple archives is frequent.⁴

APIs and interfaces

Alfresco supports industry-standard open APIs and user interfaces that improve access to enterprise content. Alfresco implements full support for Content Management Interoperability Services (CMIS), as well as providing modern RESTful APIs. Well-defined APIs mean that integration of content services with other business systems and business processes is consistent and well defined, making content an enabling technology for many business systems, including self-service web portals.

Storage management

Customer communications is one of the biggest consumers of storage in the data centre. Even with compression, customer communications archives number their storage footprint in terabytes if not petabytes. Storage tiering recognises that as information matures it needs to be transferred from short-term instantly accessible systems to less performant but significantly more cost-effective storage systems. The use of storage tiering allows companies to operationalise their information lifecycle management, the process through which information is stored according to

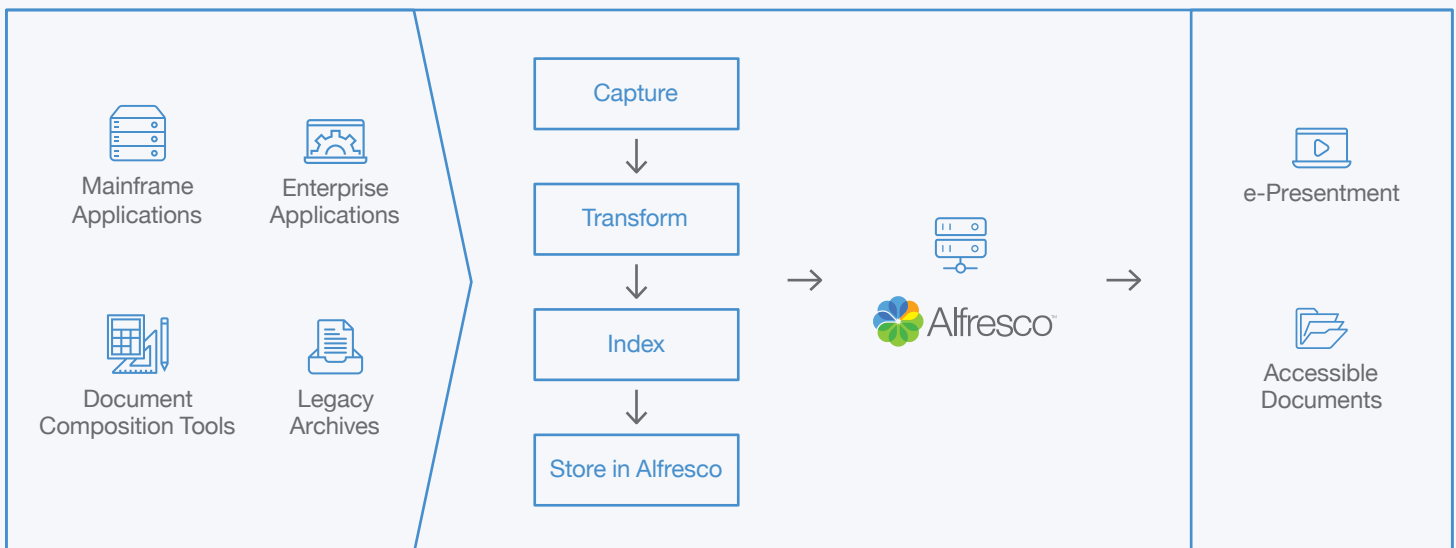
its business value, legal requirements, and the required retrieval performance.

Alfresco supports a wide range of content storage systems and provides configurable plug-ins that allow tiering and information lifecycle management. Alfresco's support for AWS, including tiered storage provision such as S3, S3 Infrequent Access, and Amazon Glacier, is an exciting prospect for customer communications archives as it permits storage to be managed in the cloud with costs accrued according to usage rather than as large capital purchases.

Ingestion

Ingestion is a crucial part of the process for loading high-volume customer communications into Alfresco. Large organisations need to load millions of documents each day, which means that ensuring the ingestion pipeline is as efficient and fast as possible is essential to maintain services to customers.

Ingestion Processing



One of the critical parts of Alfresco's customer communication archiving architecture is the use of the Crawford Technologies CCM Gateway product to manage the transformation, indexing, and loading of documents into Alfresco. The Crawford Technologies' core platform is widely deployed with legacy customer communications archives, which gives the added reassurance that Alfresco can directly support the same features found in legacy archives.

Capture from legacy systems

Legacy customer communications archives are rarely designed with their eventual replacement in mind. In fact, legacy archives suffer from their somewhat closed approach to archiving content. In particular, the use of proprietary file formats, the lack of APIs, and the paucity of unloading tools reflect the age of these products. The net result is that capturing data from these legacy archives ready for migration into Alfresco isn't always easy.

Where possible, we use documented techniques for unloading content from its source. Where techniques like these are not available, we use specialist tooling to reach into these systems and extract the content and indexes.

Alfresco's implementation partners have experience working with many of the popular legacy archives for migration and preparing their content for migration to Alfresco.

Transform and index













Crawford Technologies' core transformation modules handle print formats like AFP, Metacode, PostScript, PCL, PDF, and line data. These transformation modules are the product of over 23 years of continuous engineering and offer high-fidelity conversion between formats. Many Alfresco customers are choosing variants of PDF like PDF/A for long-term archiving since this is an internationally recognised standard with significant long-term benefits over traditional print formats.

For those choosing to store legacy print formats (as they may have done in legacy archives), Crawford provides the ability to store these in Alfresco as well. That's correct: You can continue to archive AFP with Alfresco.

INPUTS	CAPTURE	XEROX	AFP	PDF	PDF/A	PCL	PS	IJPDS	HTML5	ACCESSIBLE HTML5	IMAGE	XML
XEROX (LCDS, DJDE, METACODE)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
PS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
PDF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
AFP	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
PCL	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
IMAGES	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Load to Alfresco

Alfresco’s open-source foundations mean that it is highly adaptable to a wide range of content management use cases. For example, the most common content storage model for Alfresco is to use a one-to-one correspondence between content object and documents, which suits a wide variety of use cases and provides for fine-grained records management. With Crawford Technologies, Alfresco has developed a storage model with a one-to-many correspondence between content objects and documents, which supports high-volume archiving of document batches – a technique used in many legacy customer communications archives.

	Print	Store	Retrieve
Full Burst 20x		 Documents	 Documents
Segmented Burst 0.42x		 Segmented Batch	 Documents
No Burst 0.4x		 Large Batch	 Documents
Native No Burst 0.3x		 Large Batch	 Documents

We call these approaches “no-burst” and “full-burst” archiving respectively, and this means that our customers can trade storage efficiency with the granularity of records management. Crawford Technologies’ interface to Alfresco takes full advantage of this approach, which means that migrated content bears a close resemblance to its counterpart in the source system. The diagram shows how these different techniques can affect storage compression/inflation depending on the technique relative to an original AFP file size, i.e., full-burst archiving can increase storage by a factor of 20x.

e-presentation

e-presentation services are essential components of the processes that allow consumers to access documents via web portals and mobile devices; e-presentation can include document transformation to convert documents stored as AFP or PDF to HTML5 for display on mobile devices or a web browser. Accessible tagging is the process of formatting a document for screen readers that support the blind and partially sighted; this may need to be done in real time during e-presentation. Other operations that are performed during e-presentation may include redaction for documents that need to comply with PCI-DSS standards and digital signatures to guarantee consumers authenticity of documents.

Behind the scenes, Alfresco relies on Crawford Technologies e-presentation services to deliver these capabilities. This includes “bursting,” through which e-presentation services find individual documents contained in large batches.



Migrating to Alfresco from Legacy Customer Communications Archives

Aside from technological solutions, Alfresco has accumulated some fundamental best practice over the years that will successfully transfer your customer communications archiving workloads to our platform. Alfresco's experience combines with that of our partners Micro Strategies Inc and Crawford Technologies Inc to establish a methodology to successfully migrate legacy systems and set up day-forward archiving.

Assessing customer communications archives

It can be tempting to jump right in and begin moving things before properly scoping the size, shape, and complexity of the migration, but it is vitally important to take the time and effort to do a thorough discovery and analysis. Otherwise, there is the risk of bringing forward inefficiencies and redundancies that will make the entire migration process more cumbersome, time consuming, and prone to error.

It is not uncommon for a customer communications archive to have been in existence for a long time and it's likely that there will be some surprising and challenging content within it. Even if you're distributing your customer communications electronically, on the web, or in an e-mail, it's still common to generate a "print file" to store in an archive. So, begin your discovery by asking questions like: What types of data and documents are stored in the current archive and are they all print files? What format are those print files: AFP, Xerox Metacode, or PDF?

With a complete inventory of the files and formats contained in your archive, you will be in a much better position to properly plan the migration to Alfresco.

Sizing

The next step is to determine how much content you've got to manage from a day-forward perspective and for the migration. Day forward refers to content being archived as part of a normal daily process. This is where some crucial and strategic decisions should be made regarding exactly what gets taken on board into the new system and what is left behind or eliminated.

It is likely that communications generated on a day-forward basis will need to be archived into Alfresco. It's unlikely that you'll need to migrate every single file and every single document in the legacy archive from the start; only the most recent correspondence and high-value documents may be needed. Consider whether some information does not require immediate migration and could be removed or migrated "on demand," or in batches, over a period of time. For example, documents that exceed the regulated retention period, or customer policies that have been superseded.

You'll need to test the primary steps in the migration process, including extraction from the source, intermediate transformation and indexing, and, of course, loading into Alfresco. There will often be a pinch point in the process that needs to be scaled more than the rest. These measurements, combined with data volumes for both a day-forward basis and migration, will help size the system appropriately for migration within your desired time frame.

Bursting strategies

One of the peculiarities of customer communications archives is that they often store content in large batches. This is because the systems that generate these documents do so in overnight batch runs. When the documents were loaded into the archive, they were stored as batches and the content indexed so that the location of individual documents in the batch is known. This approach is known as “no-burst” archiving.

No-burst archiving was optimised for storage efficiency as resources, such as fonts, were stored only once for the entire batch. The corollary is so-called “full-burst” archiving, where each document is split into an individual file and stored separately. Full-burst approaches are becoming increasingly popular as storage costs decrease and the need for fine-grained records management increases with regulations such as GDPR with its “right to be forgotten.”

Fortunately, Alfresco supports both “no-burst” and “full-burst” approaches, fully supported by Crawford Technologies loading tools.

Quality and audit

Quality assurance should be the first priority of any customer communications archive migration. There are two quality factors to be considered – the quality of the archive migration and the quality of the documents themselves.

The question we need to answer is how do you guarantee that what you had to begin with is what you now have loaded into your new archive? Counting documents as they move through the extraction, transformation, and loading processes can provide a degree of assurance; however, this technique may not account for inconsistencies in the source content. Often, further objective quality measures such as randomised sampling or even semantic comparison will provide better results.

Critical to quality is document fidelity – ensuring that the exact appearance of the original document is retained in the new Alfresco archive. When transforming from one format to another, it is critically important to take the quality of the transformation into account. For example, when transforming from AFP to PDF, strategies for converting or mapping fonts will need to be taken into account.

Quality assurance should be the first priority of any customer communications archive migration by Alfresco. Only a thorough methodology and end-to-end process combined with high-quality tooling will ensure a migration will survive the inevitable quality audit.

To find out more about next-generation customer communications archiving and e-presentation with [Alfresco.com](https://www.alfresco.com), go to [crawfordtech.com](https://www.crawfordtech.com).

