



# Xplor International

## What's The Buzz About Archive Strategies

Tim Nelms, Archiving Specialist CrawfordTech

Email: [tnelms@crawfordtech.com](mailto:tnelms@crawfordtech.com)

Twitter: [timnelms](https://twitter.com/timnelms)

LinkedIn: [uk.linkedin.com/in/timnelms](https://uk.linkedin.com/in/timnelms)

Mobile: +44 7968 489105





# About Your Presenter

## Tim Nelms

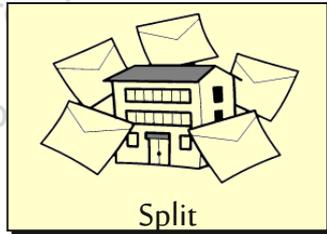
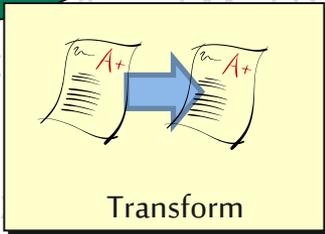
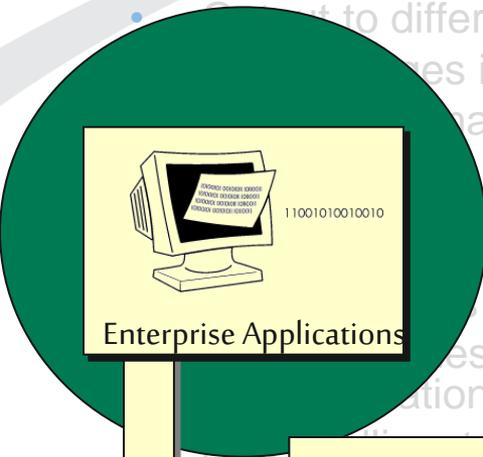
Business Development Manager for Archiving in EMEA, CrawfordTech

An IT industry veteran from the United Kingdom, specialised in software and solutions for the ECM and print sectors. Experienced in high volume content management and archiving applications since 2004 including consulting and implementing archiving strategies for some of the large enterprises in EMEA.

Currently Business Development and Alliances for Crawford Technologies focussed delivering high-volume print stream archiving solutions and transforming legacy archive infrastructures.



# CrawfordTech Print Solutions



- Connect to different devices
- Integrate with other applications
- Provide information for indexing
- Provide information for
- Provide information for multi-channel
- Provide information for
- Create print run manifests for file-based inserters & Post Office
- Automated reprint
- Track for compliance purposes
- Ensure duplicates are not printed and mailed
- Accessibility
- Alternate formats (Braille, Large Print, Audio, eText, Accessible PDF)

- eDelivery
- Electronic Envelope
- Archive & view
- Store in ECM
- Report Output Management System
- Combine multiple print files
- Extract pages and combine with other pages from other print files
- Commingling
- Transpromo
- Replace white space with 1:1 targeted messages and graphics
- Convert to color
- Printserts
- Multi Channel distribution
- Providing documents the way the customers want them



# CrawfordTech Archiving Solutions

Platform | Partners | Industry

Industry



Partner



Platform

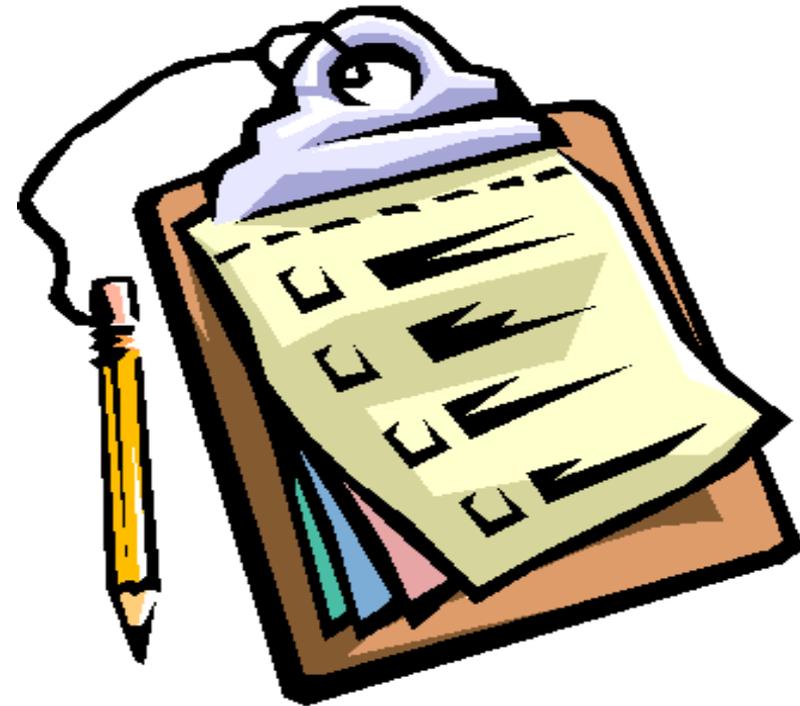
## PRO Archiver

Print Stream Archiving | Statement Archiving and Presentment | Integrated Document Archival and Retrieval Systems (IDARS) | Computer Output to Laser Disk (COLD) | Batch Workflow Automation | Enterprise Report Management (ERM) | Report Mining | Native Print Stream Encapsulation | Reprint Management | Forms Capture



# Agenda

- Why Archiving
- Archiving Evolution
- Today's Challenges
- Archiving Trends
  - Storage Formats
  - Bursting
  - Indexing
- Case Studies





# Why Document Archiving?

- Just in case....
- Disasters
- Customer service
- Business Intelligence
- Analytics
- ...when needed
- Reference
- Records
- Legal requirements
- Legal defense
- Business requirements





# Electronic archiving and its issues

- How do you preserve an electronic document today or 30 years from today?
- How do you preserve both paper-based and electronic records in a consistent format?
- How do you provide consistency in your archives?
- How do you support search and navigation?





INFORMATION

**PRODUCED AMOUNT**

**RICHNESS**

**DISPERSION**





April 16-18, 2013

Xplor  
International  
User  
Conference &







**LaserDisc**





8

7

UNIVAC

SERIES 70

UNIVAC

SERIES 70



# Information Lifecycle Management



<b>SAN</b>	<b>NAS</b>	<b>CAS</b>	<b>SAN</b>	<b>Tape</b>
<i>Block</i>	<i>File</i>	<i>Object</i>	<i>Block</i>	<i>File</i>

“The movement of information over time that recognises its business value.”

## Tiered Storage



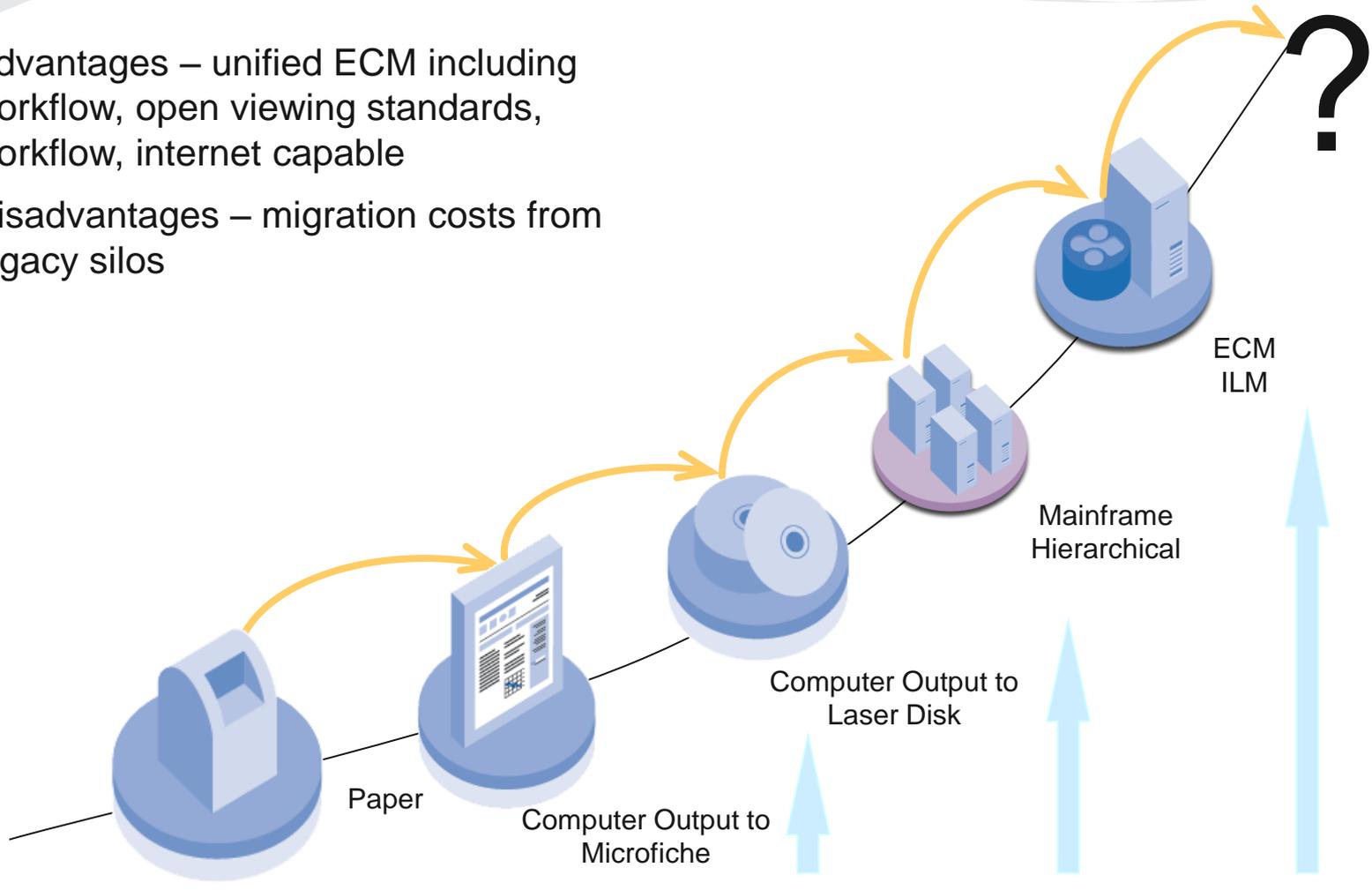


# Evolution of Archiving

Advantages – unified ECM including workflow, open viewing standards, workflow, internet capable

Disadvantages – migration costs from legacy silos

Capability



Past

Present

Future

Evolution







# Cloud Storage

- Content addressable storage
- EMC Centera
- Amazon S3
- EMC Atmos
- SNIA XAM

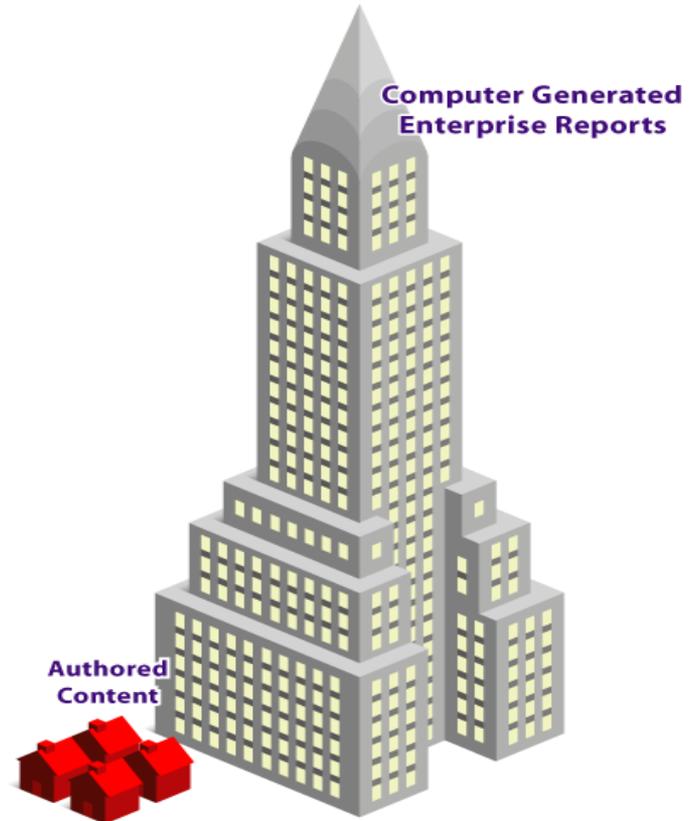




# Archive Challenges & Trends

## Challenges

- Longevity and standards
- Storage location
  - Access times
- Storage costs
  - Growth/volume
- Leveraging archived documents
  - Multiple uses, users, perspectives
  - Multi-channel delivery
- Mergers & Acquisitions
- Intelligence





# Digital Archiving Formats

- Evolved out of real-world challenges
- No formal standard (until early 1990s) to guide digital archiving
- Includes:
  - Archiving of *proprietary* formats
  - Archiving of *open standard - DeJure* formats





# Open Format Archiving

- Benefits from community driven development
- Tools, utilities and enhancements are plentiful
- Format support is more secure
- Formats are improved incrementally as they mature
- Examples of Open-standard formats for archiving
  - XML - Text - TIFF – PDF
- Archiving to an open standard or publicly available format





# Storage Formats Overview

- Original print format
  - AFP, Metacode, PCL, PS, etc.
- View formats
  - PDF, TIFF, PNG, HTML
- Long term archive formats
  - TIFF, PDF/A, XML
  - (CrawfordTech PRO Dynamic Document Archive)





	Content Accessibility, Search & Metadata	Self-Contained (Fonts, Images)	Presentation Fidelity	Platform, Application & OS Dependencies	Access to Specifications	File Size
Bitmap Formats: TIFF, BMP, JPEG	✗	✓	✓	✓	?	?
Native Formats: Office, MSWord	?	?	✗	✗	?	?
HTML	?	✗	✗	✗	✓	✗
XML	?	?	✗	✗	?	✗
PDF/A	✓	✓	✓	✓	✓	✓





# PDF –Portable Document Format

- Developed by Adobe, first published in 1993
- PDF evolved out of Adobe PostScript
- PDF introduces a standard way to communicate and present data in print and on screen

“What industries need...is a way to communicate documents across a wide variety of machine configurations, operating systems and communications networks (over time).”

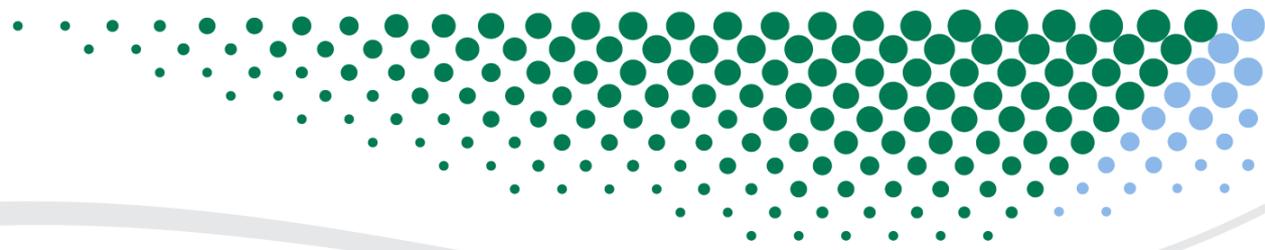
- John Warnock, co-founder of Adobe Systems, 1992





# PDF Basics

- Publicly available specification
  - Created and maintained by Adobe
- Device, Platform independent
  - Guidelines for writing, reading and printing PDF
  - PDF appears consistently across operating systems, and devices
- Third-Party Developers
  - SDK and PDF Library available, encouraging innovation
  - Extensive support from other vendors
- View and print on any platform
  - UNIX/LINUX, Mac OS, Microsoft Windows, iOS, Android
- PDF files are compact and complete, and can be shared, viewed, and printed by anyone with free Reader software
- Preserves fonts, images, graphics, and layout of any source document
- Electronic format to read any document
  - From e.g. print streams such as AFP, scanned paper and other docs into PDF
  - JPEG, MPEG, audio formats



## PDF/A - A subset of the PDF Spec

- PDF is too powerful and flexible
- Higher degree of reliability than required by the published specification
- Compatibility into the future
- Reliable migration
- Developed and maintained by an external organization





# PDF-Archive ISO TC 171/SC 2 WG 5

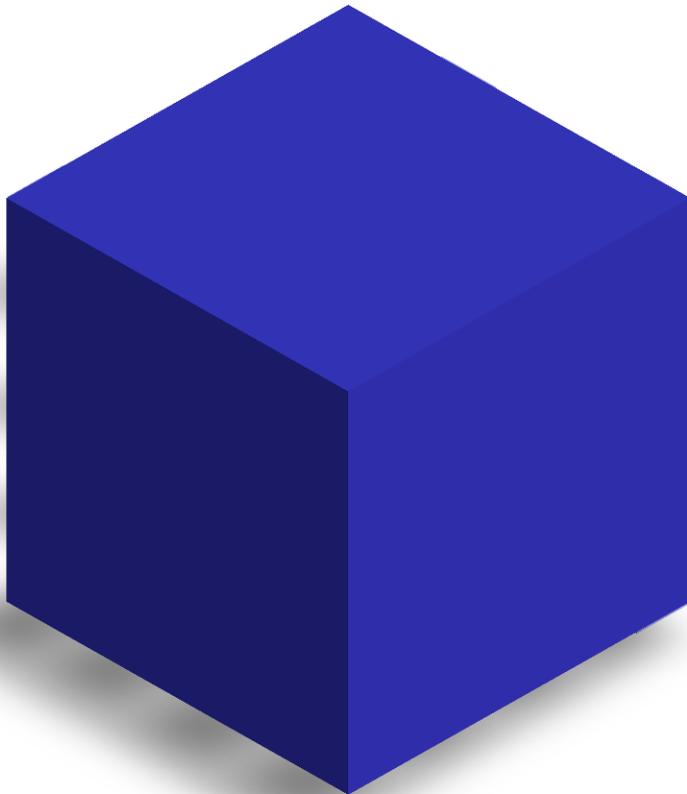
- Subset of PDF 1.4 Specification
- Specifies the use of PDF suitable for the long-term preservation of electronic documents – black and white and color compound documents as electronic data
- Final archiving format
- PDF/A alone is not an archiving solution



# Bursting Strategies

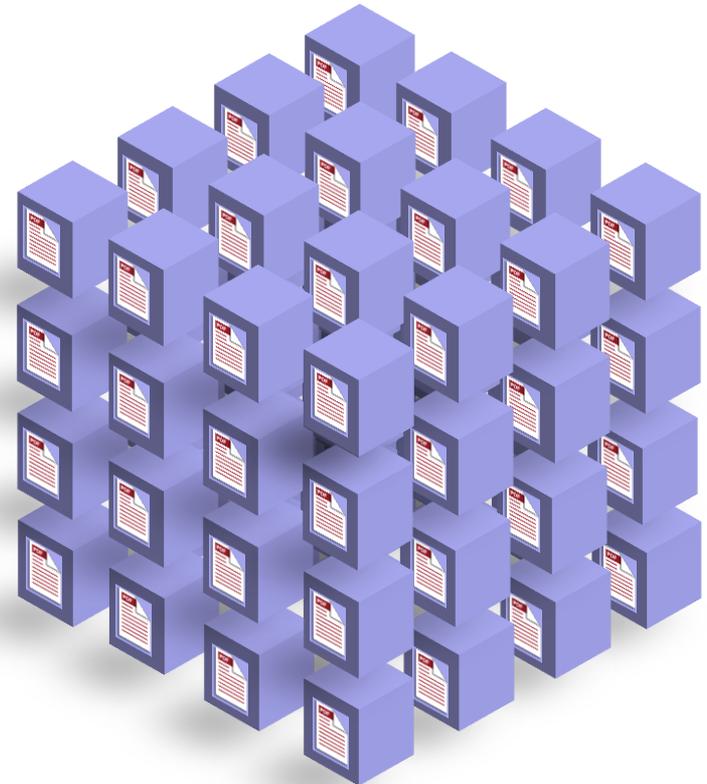
## No Burst

**All documents are retained in a single physical file, logically identified by metadata**



## Full Burst

**Metadata is used to segment documents into individual physical files; logical documents not needed**





# File Sizes

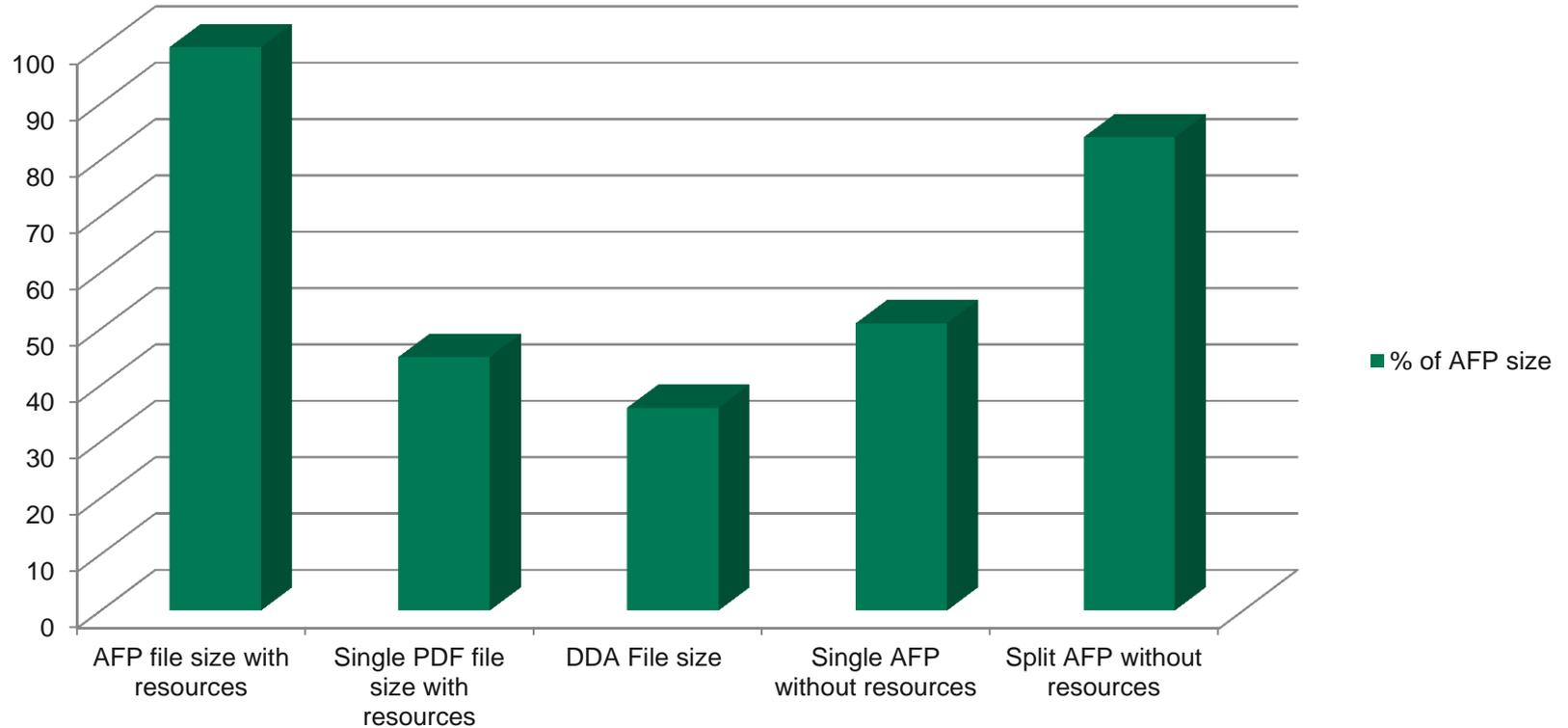


<b>Format</b>	<b>File Size</b>	<b>% of AFP size</b>
AFP file size with resources	1,372,160	100
Single PDF file size with resources	622,592	45
Split PDFs file size with resources	21,319,680	1,554
DDA File size	487,424	36
Single AFP without resources	704,512	51
Split AFP without resources	1,146,880	84

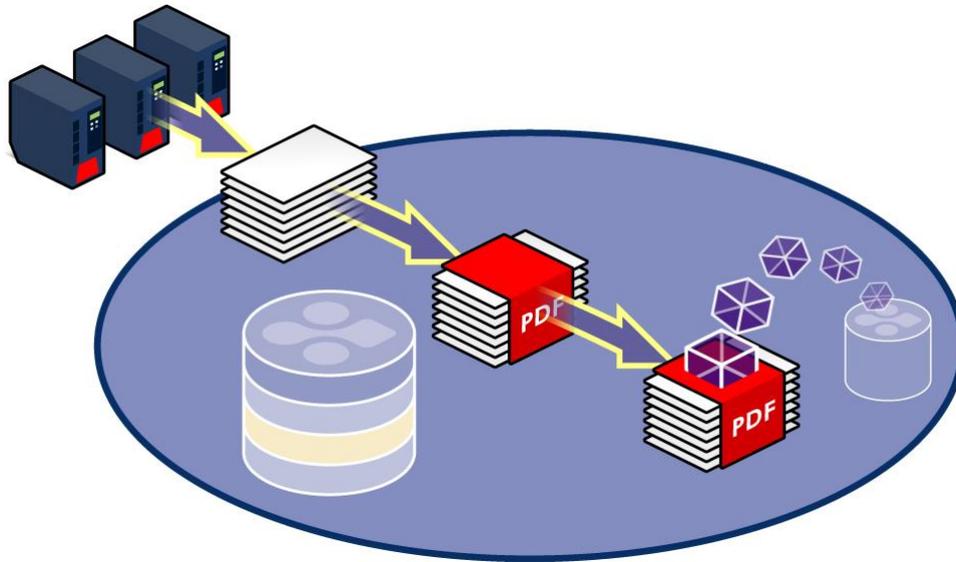




# Relative Storage Costs



# Capture Workflow



## Capture

Line of business system output as print streams

Output from legacy systems or archives for system retirement

## Normalize

Convert all input to PDF or DDA

PDF/A compliant archiving – ISO standard for digital preservation

## Index

Indexes extracted

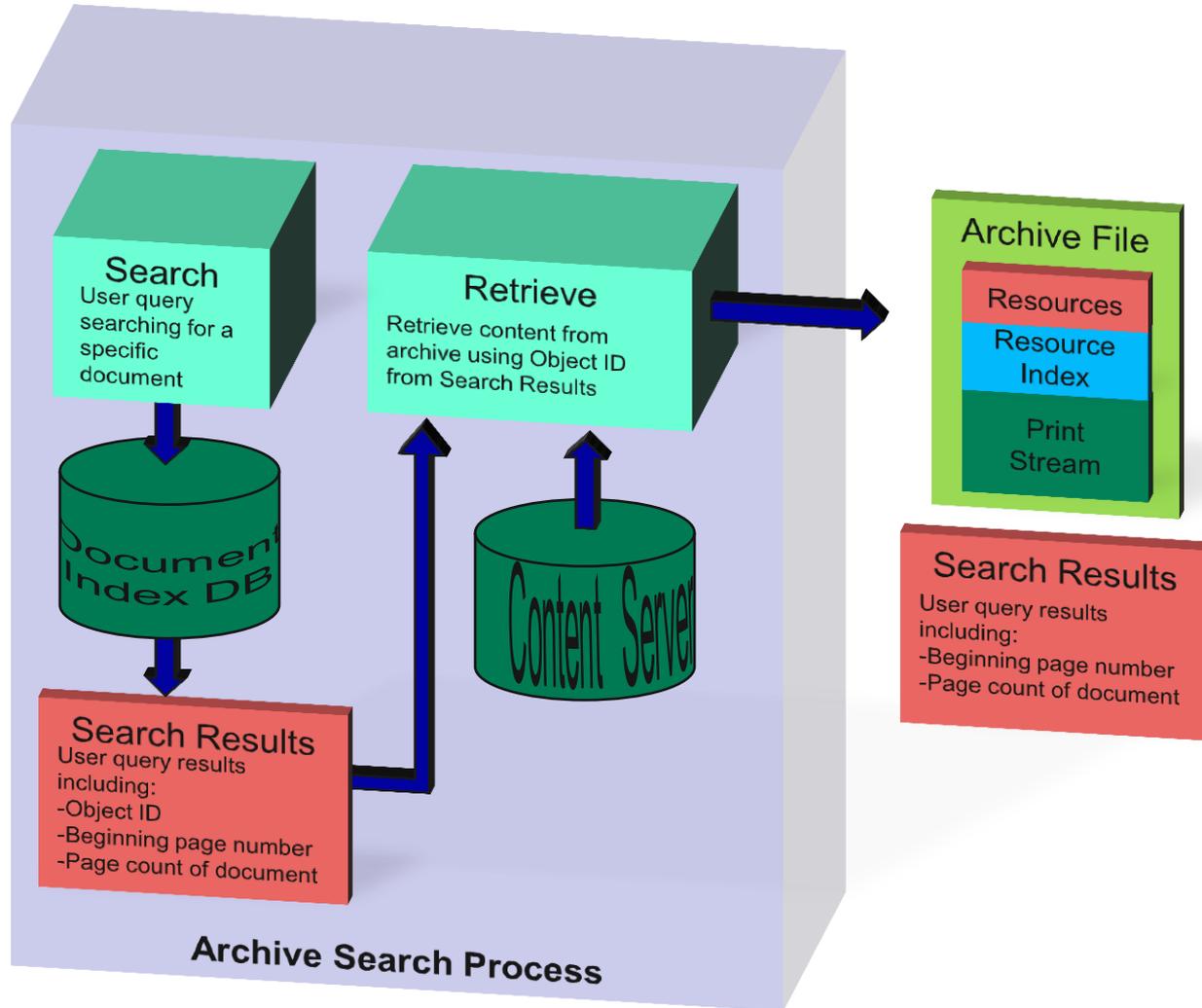
Enables classification and application assignment

## Store

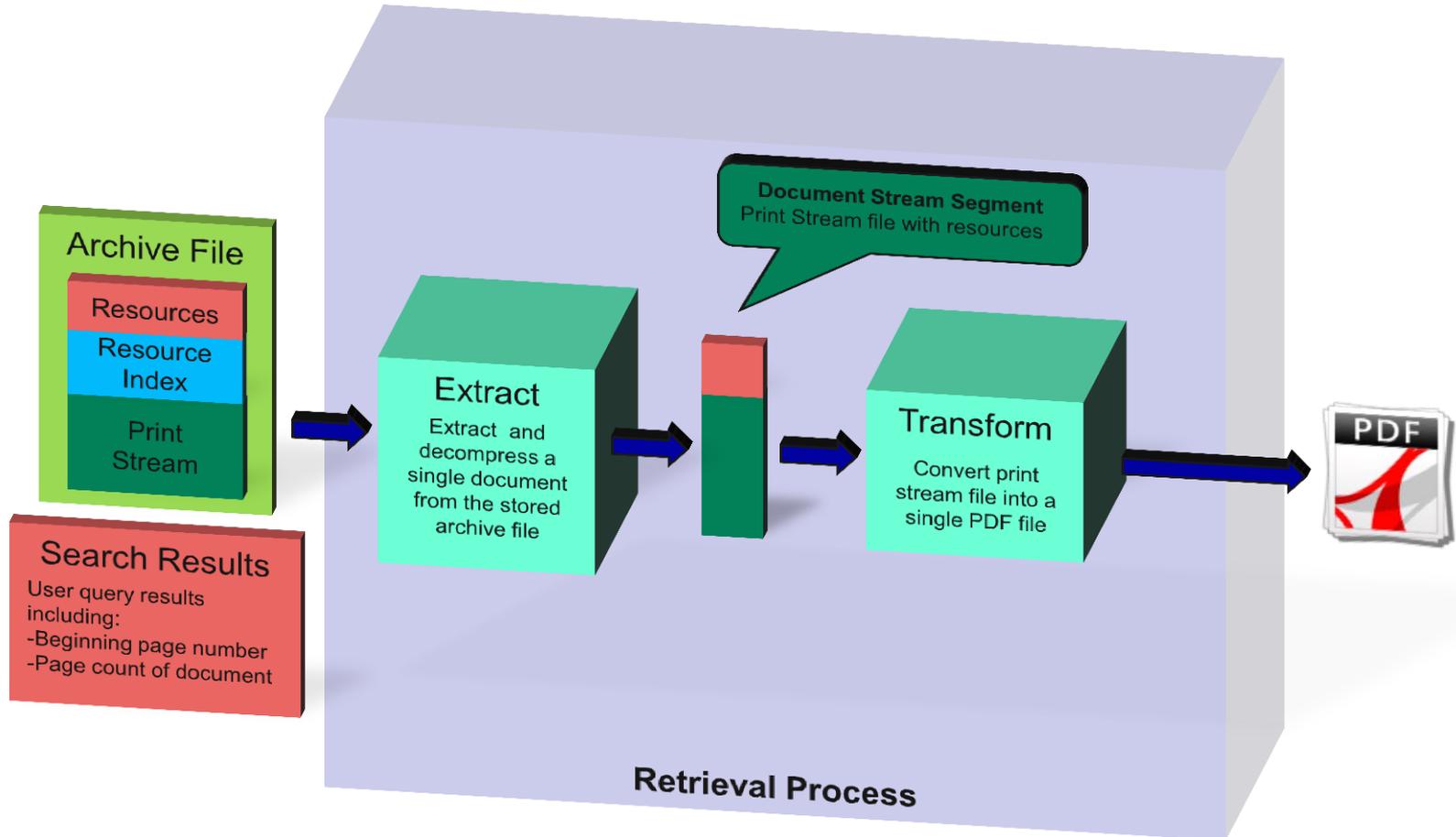
Assign ECM functions (foldering, security, workflows, lifecycle) based on pre-configured rules

Content becomes transparently available to users via existing clients

# Search Process



# Retrieval Process





# Why PDF/A and XML Matters

- Lower the cost of your archiving infrastructure
  - From multiple formats to a single standardised format
  - Open-standards ensure interoperability and independence from platforms
- Increase the value of your archived documents
  - PDF/A integrates archiving, presentation and search/navigation/retrieval
  - ISO's ratification secures long-term viability of PDF/A archives
- Mitigate Compliance Risks
  - Anticipate potential move towards PDF/A archiving guidelines
  - Provide rapid access to documents for regulatory, legal and law enforcement inquiries
  - Enable retention policies to dispose of qualifying documents





# High Volume Indexing Trends

- Vendor neutral indexing remains a challenge
- Print streams contain meta-data
- Extract meta-data into XML
- Use XML to create archive indexes for search and retrieval
- Store XML and content for index recreation during archive migrations





## Saga Insurance

Established in 1884.  
Focussed on high quality,  
value for money services for  
people 50 and over.

Products include Pet, Boat,  
Home, Motor, Motor  
Assistance, Travel, Personal  
Accident, Health Cash Plans,  
Private Medical.

Over 1000 employees  
600,000 customers – 800,000  
policies.

Producing, printing and  
posting 26,000 documents a  
day.

## Challenges

- Saga did not comply with UK Financial Services Authority regulations; the Insurance Conduct of Business guidelines (ICOB)
- Limited access to customer communications including policy documents, renewal letters, reminders and general correspondence
- The only archiving solution was print spool files, which could only be accessed by a few trained IT personnel
- Customer services staff were finding it difficult to respond to customer enquiries and could not request duplicate documents
- The costs of printing and posting 26,000 document were high

## Solution

- Saga chose EMC and CrawfordTech to provide a print stream archiving solution
- The solution converts PCL and PostScript prints streams into industry standard PDF/A and extracts customer number and policy number for indexing; the results are stored in Documentum
- Integration with insurance business systems and web portal
- The solution was implemented by EMC and CrawfordTech professional services

## Results

- Saga are now compliant with FSA ICOB regulations
- Customer service staff have instant access to customer communications which reduces call times
- Print and postage costs have reduce as customer choose digital delivery

# Insurance Conduct of Business

- A particularly demanding area of the rules is the requirement for records to be accessible to the FSA within strict timescales.
  - “The records required in ICOB must be readily accessible for inspection by the FSA.” (ICOB 2.8.2 )
  - ”A record would be "readily accessible" if it was available for inspection within two business days of the request being received.” (COB 2.8.3)
- Records may include both specific customer communications such as policy documents or generic documents where the following rules apply.
  - “Where a firm keeps standard, generic documents as records it should be able to identify which version, by date or reference number, was provided to the customer.” (ICOB 2.8.4)
- Specific guidelines to cover the keeping of electronic records and focus on the fidelity of the retained record being an accurate representation of the original document:
  - “(1) A firm may arrange for records to be kept in such form as it chooses provided the record is readily accessible for inspection by the FSA.  
(2) Where a firm chooses to maintain records in electronic form, it should take reasonable steps to ensure that:
    - (a) the electronic record accurately records the original information; and
    - (b) the electronic record cannot be subject to unauthorised or accidental alteration.” (ICOB 2.8.5)





## British Gas

British Gas (part of the Centrica group) is Britain's largest energy supplier. The company provides gas and electricity services to 15.7 million residential accounts and over one million small and medium business (SMB) and enterprise business users.

The business receives and responds to over 350,000 calls each week and generates over 270,000 letters and bills each day.

## Challenges

- British Gas operates in a highly regulated market where customer service is a key differentiator
- In order to compete with new market entrants British Gas needed to offer lower costs online tariffs
- IT systems did not provide access to customer communications including bills, letters, reminders and general correspondence
- Customer services staff were finding it difficult to respond to customer enquiries and could not request duplicate documents
- High costs for printing and posting over 70 million documents/year

## Solution

- British Gas chose EMC and CrawfordTech to provide a print stream archiving solution
- The solution converts AFP print streams into industry standard PDF/A and extracts customer number and policy number for indexing; the results are stored in Documentum
- Integration with SAP business systems and customer web portal

## Results

- Customer service staff have instant access to customer communications which reduces call times
- Customers have online access to bills and correspondence
- Print and postage costs have reduced as customers choose online only tariffs
- British Gas are now compliant with FSA ICOB regulations



## Banco Popular

Banco Popular is one of Spain's top ten Banks and the fifth largest by assets, Each month Banco Popular sends 1,000,000 documents to customers including statements, letters, reminders and statutory letters

Banco Popular wanted to demonstrate leadership in environmental matters by reducing paper consumption and postage

To meet these environmental goals Banco Popular needed to make documents accessible to customers digitally

## Challenges

- In 2008 Banco Popular took a strategic decision to improve its environmental footprint particularly concerning paper usage.
- The bank also need to lower costs in a competitive and crowded Spanish banking market
- Customer communications were identified as a key area for environmental improvements including bills, letters, reminders and general correspondence
- There were high costs for printing and posting over 12million documents/year

## Solution

- Banco Popular chose EMC and CrawfordTech to provide a print stream archiving and presentment solution
- The CrawfordTech PRO Archiver solution takes documents generated by EMC xPression and archives them in EMC Documentum as PDF/A
- Statements are made available via the E-Box online portal, which enables 150,000 customers to see, file and print out all their bank correspondence

## Results

- Banco Popular Annual Report - "A considerable volume of paper has been saved in 2009 as a result of measures taken to prevent paper from being used unnecessarily"
- Customers have online access to bills and correspondence
- Print and postage costs have reduced as customers convert to online statements



**BNP PARIBAS**

## BNP Paribas

A pan European banking group and France's 5th largest company offering, retail, private, corporate and investment banking services

BNP Paribas operates in over 85 countries with 110,000 employees

BNP inter-bank trading produces 1,400 different reports produced every month covering trading positions and branch activity

BNP had chosen to migrate away from IBM mainframe systems

## Challenges

- BNP's inter-bank trading arm produces over 1,400 reports each month
- Reports were stored on an IBM mainframe using Sungard's Store/QM product; BNP wanted to decommission the mainframe
- BNP therefore needed to migrate 10 years of history from the StoreQM archive
- Reports were archived as text but needed to be distributed and viewed more easily

## Solution

- BNP Paribas made a strategic investment in the Documentum platform for Enterprise Content Management.
- One of the first projects to benefit from this strategic investment is the replacement of an existing Sungard Store/QM system
- CrawfordTech's data mining solution supports the complex search and data mining features of the legacy system
- Reports are now stored as PDF which makes viewing easier and guarantees long term preservation

## Results

- BNP Paribas decommissioned the legacy StoreQM archive
- Bank staff have instant and secure access to reports
- Data mining allows staff to drill down into reports and access line item detail if necessary



# Summary

- Digital archiving is evolving
- Be careful when choosing formats and containers
- Insulate your archive from changes in underlying infrastructure
- Migrate archives to optimize for storage costs and performance
- Archive Silos and ECM archives have a place





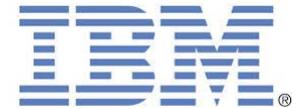
# CrawfordTech Archiving Solutions

Platform | Partners | Industry

Industry



Partner



Platform

## PRO Archiver

Print Stream Archiving | Statement Archiving and Presentment | Integrated Document Archival and Retrieval Systems (IDARS) | Computer Output to Laser Disk (COLD) | Batch Workflow Automation | Enterprise Report Management (ERM) | Report Mining | Native Print Stream Encapsulation | Reprint Management | Forms Capture





# Xplor International

## What's The Buzz About Archive Strategies

Tim Nelms, Archiving Specialist CrawfordTech

Email: [tnelms@crawfordtech.com](mailto:tnelms@crawfordtech.com)

Twitter: [timnelms](https://twitter.com/timnelms)

LinkedIn: [uk.linkedin.com/in/timnelms](https://uk.linkedin.com/in/timnelms)

Mobile: +44 7968 489105

