

Slide 1



Good afternoon and welcome to our webinar on Enterprise Records Management and Archiving. I'm Matthew Addis, CTO of Arkivum, and before that I spent over a decade working with large archives on digital archiving and preservation when I was at the University of Southampton. I'm joined by Eldin Rammell from Rammell Consulting, whom many of you already know.

The webinar will be about 30 minutes followed by Q&A and is divided into two parts.

Eldin will start us off by looking at:

Enterprise Records Management

What types of data need to be kept and why

What archiving means

What regulations apply and their requirements

How do people do it currently, or not as we'll find out!

I'll then look at:

The role of digital preservation in long-term records retention and access


The challenges and risks of trying to keep records usable for decades

What techniques and standards exist that help

How to get started

How Arkivum can help

So I'll now hand you over to Eldin



Enterprise Records Management & Archiving (Part 1)

Eldin Rammell
Rammell Consulting Limited

© Rammell Consulting Ltd 2013

If we think of what “questions and answers” come to mind in the area of enterprise records management and archiving, one of the first is; what do we actually mean by enterprise records management?

Enterprise Records Management

A systematic approach across an enterprise for the consistent lifecycle management of its records.

Typically comprises:

- Business strategy
- Technology strategy
- People strategy

© Rammell Consulting Ltd 2013 3

I define enterprise records management as a systematic approach across an enterprise for the consistent lifecycle management of its records. So, key features are enterprise-wide, consistent systems and processes and complete record lifecycle management.

Multiple IT applications, including document management systems, email, databases, forms management, web content management, workflow systems, case file management, customer relationship management. In scope if they contain records. Characteristics of records are demonstrable **authenticity** (it is what it purports to be); **reliability** (it must be a full and accurate representation of the transactions, activities or facts to which it attests); **integrity** (it must be complete and unaltered); and **usability** (it must be able to be located, retrieved, presented and interpreted). So records management systems help enterprises to manage their records in such a way that these characteristics are protected, across the record continuum, from record creation, through managed use, through to preservation and eventual disposal. Paper RM systems typically picked up the record towards the end of its life-cycle. An electronic Enterprise RM system attempts to classify and manage records based upon the record type and its value to the organisation right from its point of creation.

Enterprise RM solutions will typically not just be about technology:

Business: establishing the business case and providing value for money

Technology: deploying the right tools for the job

People: looking at the impact of the RM program across the enterprise

What types of data need to be kept?

- Documents?
- Document attributes?
- Transactions?
- Audit trails?
- Data dictionaries?

© Rammell Consulting Ltd 2013 4

Enterprise RM solutions often hold several different types of data. **Documents** are the obvious data types when we think of records management systems. Each record has **attributes** or metadata associated with it. This is usually defined to assist in the identification and retrieval of the record. But each record may also have **transactional** data associated with it, examples of which are review and approval information, data related to versioning, status of the record. Related to this are **audit** trails. This could be as simple as a record of who has accessed a record through to tracking of individual changes to the content of a record over time. And lastly, such systems often hold data **dictionaries** that help to put the records into the context of a business process. This is often linked to an external source but without the data dictionaries, the RM system ceases to function properly.

So the challenge for us is to tease apart the components of our RM systems and identify exactly what we mean when our retention policy says, for example, to keep a specific record type for 5 years. There may be competing requirements: the need to comply with a specific law or regulation may be different from our need to maintain evidential records or to support potential litigation. And our requirement to retain different types of data may actually change over the lifetime of a record.... The regulatory imperative may diminish but the litigation support requirements may increase, depending on external factors.

So when we think about long-term retention and archiving, what do we mean?

What do we mean by “archiving”?

A systematic approach to the management and preservation of inactive records that ensures their authenticity, integrity, reliability and usability is maintained.

I shall describe archiving as a systematic approach to the management and preservation of inactive records that ensures their authenticity, integrity, reliability and usability is maintained. So there are several features embedded within this definition:

- We need to identify exactly what it is that needs to be archived, understanding that Enterprise RM systems hold complex sets of related data
- Archiving should be an active, conscious activity which follows a pre-defined and documented process
- Archiving should embed best efforts to preserve all 4 characteristics of records, taking a risk-based approach

This means that the metadata that we find useful or important to manage records whilst they are active in an ERM system may be very different from the metadata that we'll need in an archive solution to demonstrate authenticity, integrity, reliability and usability.

What regulations apply?

- OECD Advisory Document “Establishing and control of an archive that operates in compliance with the principles of GLP”
- GLP
- GCP
- GMP
- GPvP
- GDP.....

I’m often asked what regulations apply to archiving? Well there are very few regulations that deal **specifically** with archiving. In fact the only one that comes to mind is the OECD Advisory Document “Establishing and control of an archive that operates in compliance with the principles of GLP”. In other cases, we need to refer to existing regulations that cover the scope of activities to which the records relate. And the degree of guidance that these predicate regulations offer varies. In some cases there is almost no acknowledgement of electronic archiving and electronic records.... in which case you need to think very carefully about how the regulations are to be interpreted in our electronic worlds. In other cases, however, there are very specific references to the requirements for electronic records. So lets come on to some of those requirements:

What are the key requirements?

Media independent:

- Named archivist (GLP/GCP)
- Clear ownership
- Review prior to archiving
- Tracking systems
- Appropriate storage

Firstly, there are some archiving requirements that apply whether we're talking about electronic or hard-copy:

[talk through bullets]

But there are also requirements in the applicable regulations that relate specifically to electronic records:

What are the key requirements?

- Appropriate media
- More than 1 copy
- Restricted/controlled access
- Prevent unauthorised change
- Guaranteed future access
- Protection against h/w deterioration
- Periodic testing of h/w
- Validated transfer processes

[talk through bullets]

Guaranteed future access – should be described in **written procedures** how this will operate

This is just an abbreviated list but as we think about our Enterprise Records Management systems, it soon becomes clear that compliance with some of these requirements will be a challenge. So, what approaches are being taken?

What approaches are currently used?

- No strategy!
- Delay strategy (“we’ll decide soon...”)
- Print to paper (yes, really!)
- Portable storage media (e.g. DVD)
- **Archive tools as part of a digital preservation strategy**

Surprisingly, many organisations do indeed have no strategy for archiving electronic records. In these organisations, inactive electronic records are maintained in live IT applications as the storage and backup requirements and costs get larger and larger, year on year. Other organisations have identified the problem but have not yet decided on the solution. They are either putting off the decision-making or just haven’t yet got to that point in time.

Incredibly, there are organisations that use hard-copy archives as their corporate solution, even for records that are born-digital and otherwise would not exist in paper form. Increasingly, regulatory authorities around the world are not accepting this as a viable solution, especially as important metadata is often lost when the records are printed down.

And for some applications, archiving on DVD is seen as an acceptable solution, but there are many problems associated with this approach, not least of which is the potential rapid deterioration of the storage media.

And so the ideal approach is to identify specific IT tools that are “fit for purpose” for electronic archiving as part of an overall digital preservation strategy. But it is critical that the tools are indeed **fit for purpose**, based on the records that are being supported and our record-keeping requirements.

And at this point, I’ll hand over to Matthew Addis to continue on the theme of digital preservation.

Enterprise Records Management & Archiving (Part 2)

Matthew Addis, Arkivum Ltd

Key Questions

- What is digital preservation?
- What risks does digital preservation counter?
- What about digital preservation standards?
- How can digital preservation be applied to records management?
- How can Arkivum Help?

So in this part of the webinar we'll look at the role of digital preservation in long-term archiving of records and associated data.

We'll start by looking at what digital preservation is, what the risks are to records when trying to retain them for the long-term, the use of digital preservation standards as a guideline on how to tackle the problem, how digital preservation techniques can be applied to records retention, and then finally how Arkivum's solution can help.



So what is digital preservation and how does it relate to records management?

Digital Preservation

"The series of managed activities necessary to ensure continued access to digital materials for as long as necessary."
Digital Preservation Coalition

"Digital information lasts forever - or five years, whichever comes first."
Jeff Rothenberg

© Arkivum Ltd 2013

ARKIVUM
POLICY CENTRE

There are many definitions of Digital Preservation, which are broadly the same, and all are concerned with the ability to access and use digital content over time. The definition here is from the Digital Preservation Coalition [1]

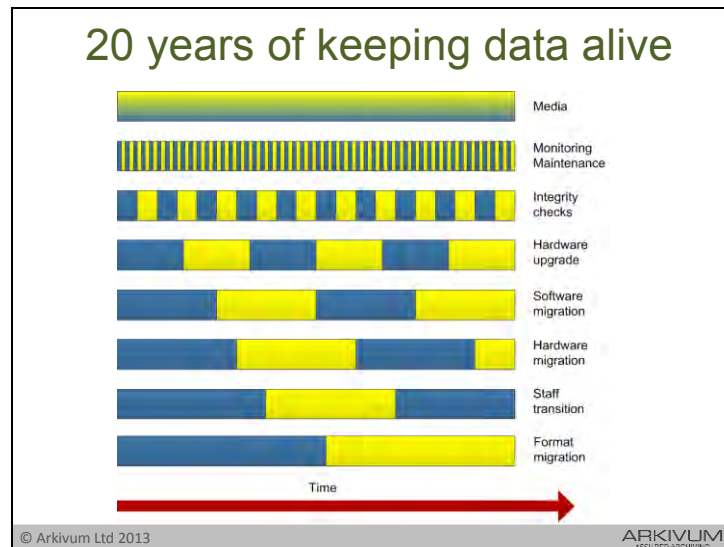
It's the bit about being able to use content that's key to preservation – it's not just storing something and showing that it's authentic and hasn't been modified, it's about being able to read and understand that content in the future.

But perhaps my favourite alternative definition comes from Jeff Rothenberg, which states that the natural lifetime of digital information is just 5 years [2]

If you want content to live longer then you need to make active steps to keep it alive, and that's what digital preservation is all about.

[1] <http://www.dpconline.org/advice/preservationhandbook/introduction/definietions-and-concepts?q=definitions>

[2] <http://www.clir.org/pubs/archives/ensuring.pdf>

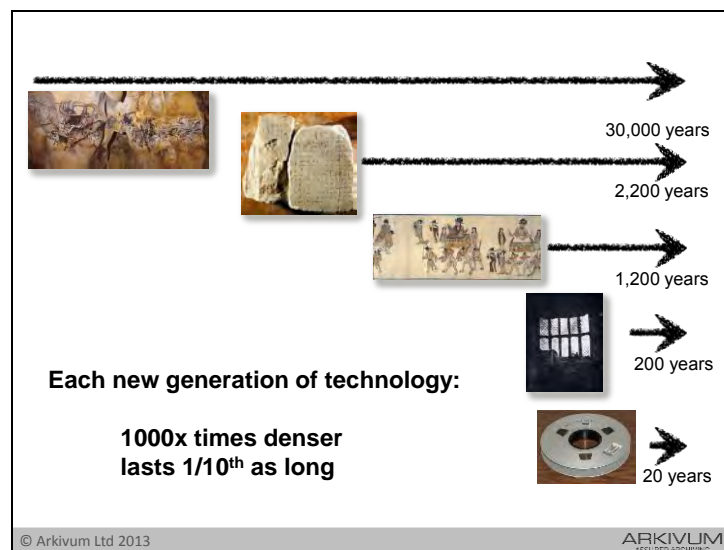


What Jeff was referring to is obsolescence that happens in the digital world and the need to be pro-active in countering it.

This picture shows just some of the things that will happen over 20 years of trying to retain data.

In the diagram, a change from blue to yellow is when something happens that has to be managed. In a growing archive, adding or replacing media, e.g. tapes or discs, can be a daily process, so is effectively continual. The archive system needs regular monitoring and maintenance, which might mean monthly checks and updates. Data integrity needs to be actively verified, for example annual retrievals and integrity tests. Then comes obsolescence of hardware and software, meaning refreshes or upgrades that will typically be 3 – 5 years, for example servers, operating systems, application software. In addition to technical change in the archive system is managing staff transitions of those who run the system, for example support staff and administrators. Even the format of the data being held may need to change because applications that read the data become obsolete or get upgraded and are no longer backwards compatible.

The point is that there is a continual series of changes that need to be managed.



And these changes happen at all levels.

Technology for storing content is just one example.

As humanity generates every more data, technological advances are made to store and access it.

Advances get faster and storage gets cheaper, but the new technologies created last for less time – they become obsolete ever sooner.

We're now down to the level where almost every cost effective storage technology lasts less than 5 years. It could be data tape, it could be hard drives, it could be solid state. Nothing lasts more than 5 years anymore - not necessarily because the media itself decays, but because the drives or systems that read that media become obsolete.

Obsolescence is the main enemy to long-term data retention and happens on worryingly short timescales.

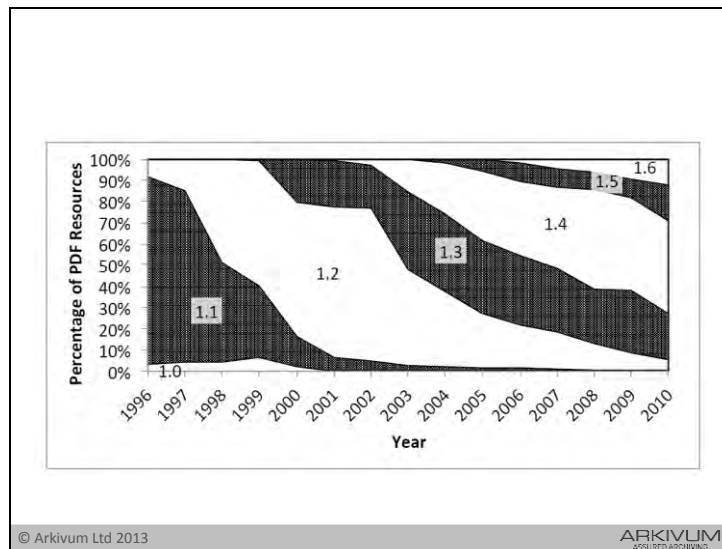
Just to illustrate this, a guy called Andrew Brown was in the news late last year. He asked his local hospital for a copy of an echo cardiogram that was performed on him in 2004. The BBC [1] reported that the Worcestershire Acute Hospitals NHS Trust said it would cost £2000 for him to be given a copy of his data. The Register reported that the trust said this "was not a cost-effective use of public money" [2]. The hospital does have his echo cardiogram data, which is stored on Magneto-Optical disk, but the hospital no longer has a drive that can read these disks as they have subsequently installed a new archive system [3]. Their supplier apparently said that they didn't stock the drive anymore as it was no longer in production

and they would have to ship a drive in from the United States if the Hospital wanted to read the data. That's technical obsolescence in action – and in just over 6 years. This sort of thing really does happen. It illustrates why you need to think carefully whether 'long-lived' media is the right way forward. It's tempting to think that using specialist storage media that's designed to last for decades or centuries is the answer, including magneto-optical disks and other forms of 'archival grade' storage. But this technology often addresses a niche market, which means it can be harder for the companies who develop and sell it to make a sustainable business. The storage media might last for decades, but the companies who make it might not, or they are forced to move on to develop something new.

[1]<http://www.bbc.co.uk/news/uk-england-hereford-worcester-20235193>

[2]http://www.theregister.co.uk/2012/11/08/nhs_scan_2k/

[3]<http://www.whatdotheyknow.com/request/94658/response/237590/attach/2/attachme nt.pdf>



The same problem applies to the applications and formats used to store digital content.

This picture shows the use of different versions of PDF as seen by the British Library when harvesting web content in the UK [1] from 1996 to 2010.

New versions of PDF are available every few years, they get widespread adoption, but then rapidly get superseded by new versions so their use tails off, although it does take a long time to die completely.

For PDF maybe this isn't so much of an issue as backwards compatibility is good.

But what about all the other types of content, for example correspondence, supporting data from laboratory systems, patient scans, etc. That's the stuff to really worry about.

There was an initiative set up November last year on file formats as a crowd-sourcing exercise to list file formats and their specifications. This already contains 140 formats for biological, biomedical and medical imaging data [2] I bet that's the tip of the iceberg. How many of those will be obsolete in the next 10 years?

[1] http://www.scape-project.eu/wp-content/uploads/2012/11/iPres2012_Formats-over-time.pdf

[2] http://fileformats.archiveteam.org/wiki/Scientific_Data_formats

File formats and migration

Droid

DROID (Digital Record and Object
Identification)

The **technical registry**

PRONOM

Media type	File formats	Preservation format(s)	Access format(s)	Normalization tool
Audio	AC3, AIF, MP3, WAV, WMA	WAVE (LPCM)	MP3	FFmpeg
Email	PST	MBOX	MBOX	realtext
Email	Mails*	Original format	MBOX	md2ms.py
Office Open XML	DOCX, PPTX, XLBX	Original format	PDF for PPTX	OpenOffice
Plain text	TXT	Original format	Original format	None
Portable Document Format	PDF	PDF/A	Original format	Ghostscript
Presentation files	PPT	Original format	PDF	OpenOffice
Raster images	BMP, GIF, JPG, JFIF, PCT, PNG*, PSD, TIFF, TGA	Uncompressed TIFF	JPEG	ImageMagick
Rare camera files/Digital Negative format*	JFR, ARW, CR2, CRW, DCR, DNG, ERF, KDC, MRW, NEF, ORF, PEF, RAF, RAW, X3F	Original format	JPEG	ImageMagick/LFRaw
Spreadsheets	XLS	Original format	Original format	None
Vector images	AI, EPS, SVG	SVG	PDF	Inkscape
Videos	AVI, FLV, MOV, MPEG-1, MPEG-2, MPEG-4, SWF, VMOV	FFV1/LPCM in MKV	MPEG-1	FFmpeg
Word processing files	DOC, XPS, RTF	• ODF (ODP and RTF) • Original format (DOC)	PDF	OpenOffice

© Arkivum Ltd 2013

ARKIVUM
FOUNDED 2013

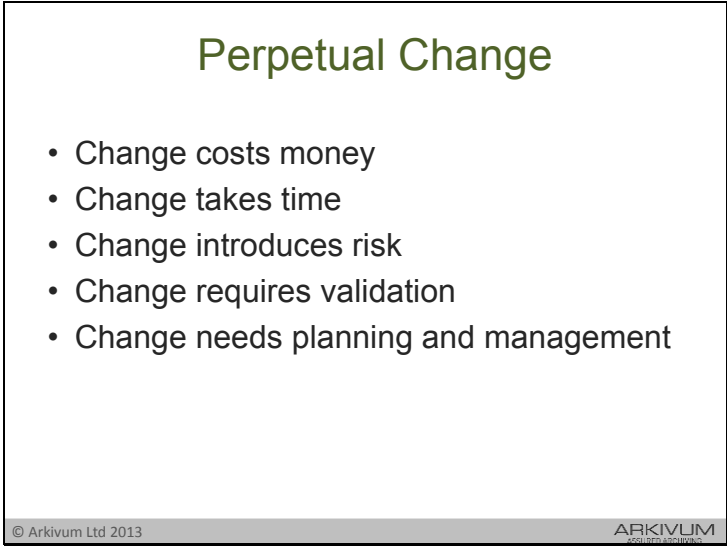
In the library and archive community this is a recognized problem. For example, there are format identification tools and services such as PRONOM and DROID from the national archive [1], and guidelines on what formats to choose for long-term accessibility. But support for scientific data formats is very limited.

So the most common strategy is migration to an open, well documented and long-lived format - often when the data is first archived, but sometimes at a later date if the format of the data isn't immediately of concern. This is often PDF-A for documents. But again the challenge is all that other supporting digital content.

The approach is typically to migrate multiple data formats into one or more canonical forms, as shown in this table for the Archivematica [2] tool. The key thing is that the format for long-term preservation isn't typically the format data was first created in, and neither is it the one best suited for access and use. The objective is to think about longevity first and foremost.

[1] <http://www.nationalarchives.gov.uk/information-management/our-services/dc-file-profiling-tool.htm>

[2] https://www.archivematica.org/wiki/Main_Page



Perpetual Change

- Change costs money
- Change takes time
- Change introduces risk
- Change requires validation
- Change needs planning and management

© Arkivum Ltd 2013 ARKIVUM
POLICY CENTRE

So what we have is a picture of digital preservation being about continual change.

It's about those blue to yellow transitions. Each one is a change. And:

Change costs money

Change takes time

Change introduces risk

Change requires validation


Change needs planning and management



And the point about change = risk is the one I want to pick up on in the next question.

Risks

- Technical obsolescence, e.g. formats and apps
- Hardware failures, e.g. digital storage systems
- Loss of staff, e.g. skilled archive and IT staff
- Insufficient budget, e.g. storage too expensive
- Accidental loss, e.g. human error
- Selection, e.g. don't retain what you should
- Stakeholders, e.g. retention no longer a priority
- Underestimation of resources or effort
- Fire, flood, meteors, aliens...



© Arkivum Ltd 2013 ARKIVUM
POLICY CENTRE

So you'll have already got the feeling that many things that can happen that cause loss of content. Some are technical, e.g. storage failures, but some are to do with people, processes, planning and unforeseen events [1]

EDRM systems themselves can introduce risks, especially if they are not designed with long-term preservation in mind. They can be great ways to handle records when they are first created and structured, but are they the best way to store records for the long-term?

What's needed is a thorough risk assessment activity that considers all aspects of data retention, including how well existing systems and processes stand up.

[1]

https://prestoprimews.ina.fr/public/deliverables/PP_WP3_ID3.2.1_ThreatsMassStorage_R0_v1.00.pdf

Risk Assessment and Compliance	
Integrity	proving data hasn't changed
Authenticity	proving where the data came from
Availability	access when you need it
Confidentiality	control over who can read it
Possession	knowing who has it
Utility	knowing you can still use it
Provenance	chain of custody
<div>© Arkivum Ltd 2013</div> <div>ARKIVUM POLICY CENTRAL</div>	

The key to doing this is to align the risk assessment with existing standards or regulations:

If you look across the range of regulations, be it 21CFRpart11, EU Annex 11 on GMP, UK MHRA requirements on GCP, the same common elements often come up: integrity, authenticity, confidentiality and usability.

These then lead to 7 questions you can use to frame a risk assessment from a preservation perspective.

Risk Management

- Do nothing
- Do the wrong thing
- Do it in-house
- Use a service provider



- ISO27001 Information Security Management
- ISO16363 Trusted Digital Repositories

© Arkivum Ltd 2013 ARKIVUM

The biggest risk for preservation is of course to do nothing - indecision and delayed action – or maybe assuming that just leaving data in an EDM system is OK.

Then if you do something then there's the question of making the wrong choice.

There's also the issue of whether you do it yourself, which requires the necessary in-house skills and expertise, or whether you outsource to a service provider to take advantage of the knowledge and infrastructure that they may have but you don't.

It's all about being informed and managing risk, including managing the risk of using a third-party to provide some or all of the solution.

Beyond regulations, there are two key standards to be aware of: ISO27001 [1] for information security management and more recently ISO 16363 [2] which used to be: TRAC trusted repository audit criteria.

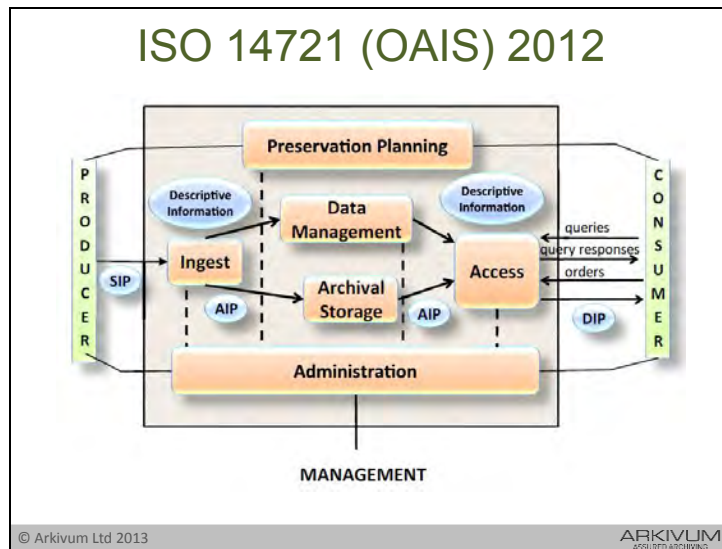
The value is that both standards provide a useful checklist of what to look for in solutions or what to do if you're building your own.

[1] <http://www.27001-online.com/>

[2] <http://public.ccsds.org/publications/archive/652x0m1.pdf>



So that mention of standards brings me to the next question.



The first standard to mention is OAIS [1] – Open Archive Information System – from the Consultative Committee for Space Data Systems, or other wise known as ‘brought to you by NASA’, although for the recent iteration many more space agencies have been involved, so it’s a major initiative.

OAIS provides a framework for long-term retention and access to digital content using several key principles, for example:

- Ensuring you capture all data and metadata up-front according to a submission agreement to ensure you have everything needed in the long-term, which is important as it breaks the dependency on the original producers of the content to understand what it means and how to interpret it.
- The need to support what OAIS calls a Designated Community, which are the people who will need to use the content, for example regulators or internal reuse. This is important and the designated community defines what key characteristics need to be preserved in digital content.
- Then in the middle is the internal planning and managing of the storage and preservation actions so that the designated community can be served and the key characteristics of the content that are important to this community are maintained.

This approach is a bit different to EDRM systems which, by and large, focus on the earlier part of the lifecycle where content is first created, not when it subsequently needs to be preserved. So what OAIS provides is a framework for assessing whether an EDRM has the right features to support digital preservation.

[1] <http://public.ccsds.org/publications/archive/650x0m2.pdf>



Then comes the issue of metadata, which means metadata needs to know first how to do the preservation, second that preservation has been followed, and thirdly demonstrate this in a way that is auditable.

There are standards for records management [6], for example the ISO 23081 series, which builds on ISO 15489 and addresses metadata for records management. And there's MoReq2010 [1] which is the latest in a series of requirements frameworks for records management systems. These build on ISO 15489, which is a records management standard that requires records to have the characteristics of authenticity, reliability, integrity, and usability.

But there are also standards for metadata from the digital library and digital preservation community, which include:

PREMIS, which is a standard for metadata associated with the preservation of digital objects and might, for example, contain checksums, signatures, format descriptions, file structures, and a list of events associated with preservation such as migrations.

And

METS, which is a standard for encoding various types of metadata into a form that can easily be stored or exchanged. Metadata put into a METS wrapper might include administrative metadata, e.g. PREMIS metadata, but also descriptive metadata of the content of a digital object, and how the components of an object are structured, e.g. a set of files.

These have a lot in common, e.g. they have the concept of events that need to be recorded that form part of the lifecycle of an object or record, e.g. when it was created, migrated, disposed of etc. And they have the concept of fixity, authenticity, usability etc. that align with regulatory requirements.

So, whilst there are:

- increasingly well developed standards for digital objects themselves, e.g. eTMF[5] or eCTD [4], and
- regulations that require the retention of these objects in a way that can assert integrity, confidentiality, authenticity and accessibility,
- what the metadata standards do is to allow the preservation of these objects to be performed, managed and recorded a lot more effectively

And most importantly these standards allow it to be done in a vendor neutral way, i.e. they do not assume any specific products, tools or infrastructure. Indeed the opposite is true – preservation metadata can be structured and recorded using simple XML formats and stored in files on a file system.

[1] <http://moreq2010.eu/>

[2] <http://www.loc.gov/standards/premis/index.html>

[3] <http://www.loc.gov/standards/mets/>

[4] http://estri.ich.org/eCTD/eCTD_Specification_v3_2_2.pdf

[5] <http://www.etmf.org/>

[6] http://www.armaedfoundation.org/pdfs/V_Jones_RIMStandards_Update2012.pdf



But how do you know your solution is a good one?

And how to you assess the risks and identify if good practice is being followed?

This is where standards such as ISO16363 come in [2]. This is a standard for assessing whether a digital repository is trustworthy, not just in the sense of the measures it uses to protect digital objects, but also its economic sustainability, business continuity plans, disaster recovery, legal and contractual arrangements with its users and so on.

The TDR standard also builds on work in a project called DRAMBORA [1] amongst others, and DRAMBORA developed a risk assessment based approach.

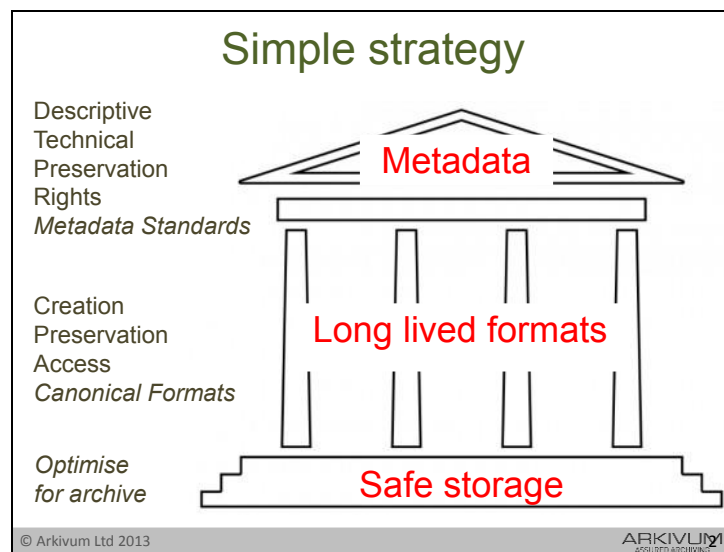
So, TDR and Drambora provide that risk assessment and audit piece of the puzzle.

[1] <http://www.repositoryaudit.eu/>

[2] <http://www.iso16363.org/>



So all this might sound a bit complicated – digital preservation requiring a whole pile of complex standards and best practice. But these can be applied quite simply and effectively.



A simple approach to get going is to build a 'preservation house'. Start at the bottom and build up.

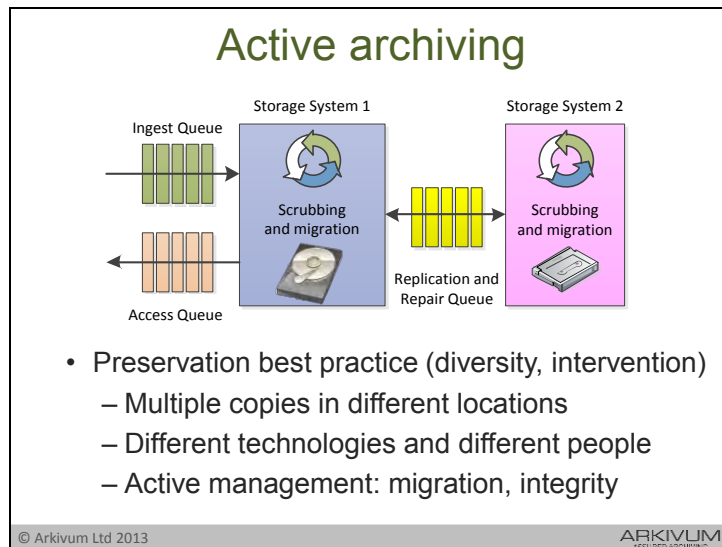
The key thing is solid foundations – storage and storage management that won't lose your data

Then comes file formats – the approach here is to choose a long-lived format if possible, e.g. PDF-A

Finally, on top of the house is the roof, which means metadata to describe what data is being held and allows it to be found again in the future. Metadata can generally be split into four types:

Descriptive metadata that says what's in the file,
Technical metadata that says what the format is,
Preservation metadata that says what to do, or has been done, to the data to keep it accessible, e.g. migration
Rights metadata that says who owns or can use the data, e.g. IPR

These should all be captured in an open format, e.g. XML, ideally using a standard, e.g. METS and PREMIS.




So what do you do to get started at the safe storage level – the foundation of the preservation house?

- You make multiple copies of the data and you keep them in different locations.
- You use diverse technologies to spread the risk of failures, which effectively means not all eggs in one basket
- And you actively manage these copies by (a) migrating to new storage or formats to address obsolescence and (b) by regularly checking and repairing any loss of data integrity (which is why having multiple copies is so important – if there is a problem with one of the copies then it can be replaced by replicating one of the other good copies)

This foundation is often ignored by EDRM and even if it is put in place, then there typically isn't integration so you know at the management level whether it's functioning correctly.

Burnline

- Metadata and data on a file system
- Open standards and formats
- Drive down costs and risks



© Arkivum Ltd 2013 ARKIVUM
POLICY CENTRE

The other simple strategy is to support what's called a 'burnline'.

This is a term I've borrowed from Neil Jeffries at the Bodlean Library and refers to safeguarding research data.

The idea is that if there is some form of disaster, the equivalent of a burning building, then the only thing that you need is the storage system used to hold the data. This is because on storage you put a complete record of all the data and associated metadata – in addition to, or instead of, using complex records management systems, databases, sharepoints or whatever else. This data and metadata is held in open formats and using open standards. Everything can be rebuilt from that if necessary.

The point is that this approach is just good preservation practice and provides extra assurance in a way that is:

Simple.

Removes complexity.

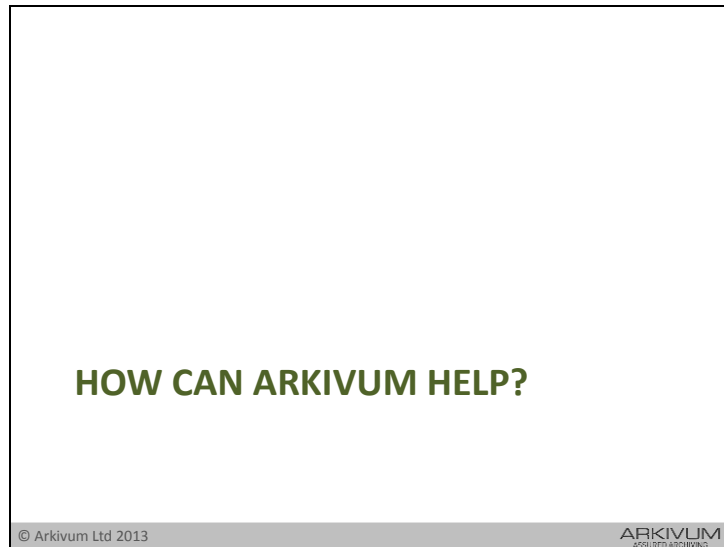
Removes dependencies, including lock-in to keeping everything in an EDRM.

Is more manageable over the long-term.

Lowers the risks

Keeps costs down and predictable

Slide 31



So how does Arkivum fit into the picture.

A presentation slide for Arkivum. At the top, the word "Arkivum" is written in a green, sans-serif font. Below it is a bulleted list of five points: "Online data archiving as a service", "Spin-out of University of Southampton", "Decade of know-how working with archives", "Safe, secure, accessible data storage", and "Designed from ground-up for retention and access". Below the list is a dark grey rectangular box containing the Arkivum logo on the left and the text "100% data integrity guarantee" and "Keep your data safe & secure forever" on the right. The logo consists of the word "ARKIVUM" in blue and yellow, with "ASSURED ARCHIVING" in smaller blue text below it. At the bottom of the slide, there is a thin grey bar with "© Arkivum Ltd 2013" on the left and the Arkivum logo on the right.

Arkivum

- Online data archiving as a service
- Spin-out of University of Southampton
- Decade of know-how working with archives
- Safe, secure, accessible data storage
- Designed from ground-up for retention and access

ARKIVUM
ASSURED ARCHIVING

100% data integrity guarantee
Keep your data safe & secure forever

© Arkivum Ltd 2013

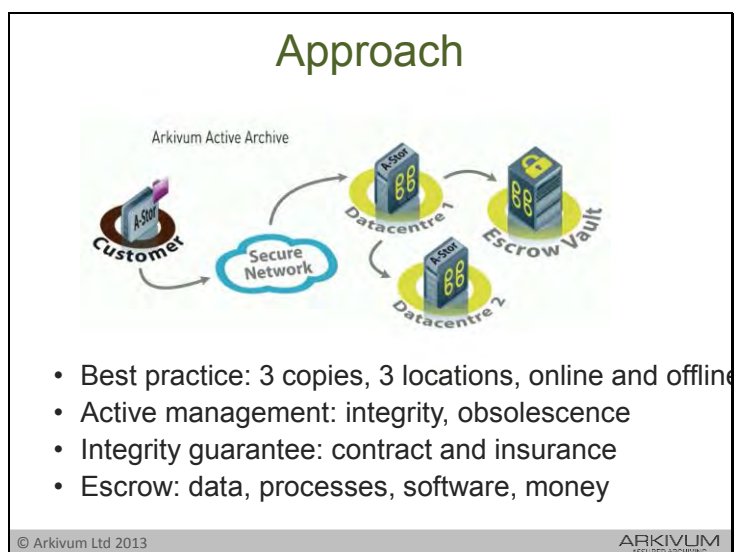
ARKIVUM
FOUNDED 2011

Arkivum provides online data archiving as a service for those organisations that need to keep data for the long-term for compliance or reuse. We have customers in healthcare, life sciences, energy and voice call recording to name but a few.

The company was founded in 2011 as a spin out of the University of Southampton and is based on expertise the founders have from working on digital preservation, for example from working with national broadcasters and archives across Europe.

Put simply, we provide the foundation of the preservation house.

And we are a great way of providing a burnline.



The approach we take is to follow data preservation best practice. Three copies of customer data are held in three known UK locations. We use checksums to actively manage data integrity through regular checks and we do regular media and infrastructure migrations to counter obsolescence and to ensure costs remain low.

The solution is provided as a service, but can also be installed and operated in-house if required.

Basically, we take care of that 'blue and yellow stuff' in that diagram I showed earlier. And it's our skilled and dedicated staff that do this that you're also getting access to, not just the infrastructure we use.


We are certified to ISO27001, which means Arkivum has been externally audited for the integrity, confidentiality and availability of data assets in our possession. We follow the OASIS model. We have done our own DRAMBORA risk assessment and we apply ISO16363 principles. We haven't been audited as a Trusted Digital Repository just yet, but that's because accreditation for auditors is still being set up. What we have done is all the hard work on following digital preservation standards and this takes some of the burden off our customers.

It also means we can offer a guarantee of data integrity. All data returned from our service is always bit-for-bit identical to the data the customer supplied, with no restrictions on time or volume. The guarantee is backed by insurance and is included in our SLA.

Finally, one of those copies of customer data is held at a third-party site with a three way agreement in place with us, the third-party and our customers so that if they want to leave our service, or if we can no longer provide the service we agreed, then the customer gets direct access to a complete copy of their data on media that they can take away. This is part of a wider escrow model that includes the software and processes we use to run the service as well as ring-fencing of money for fully paid up long-term retention contracts.

Pricing

- PAYG or Paid-Up for 5,10 or 25 year
- No ingress or egress charges
- Migrations and refreshes included
- Audits and certification included
- Escrow copy included, no exit costs



© Arkivum Ltd 2013 ARKIVUM
POLICY CENTRAL

The pricing model for the service is simple.

We support both opex, i.e. PAYG, and capex, i.e Paid-Up, models. So, for example, we can offer retention for 10 years for a fixed up front cost. And, unlike other storage service providers, there's no charges for getting data in and out of our service.

And, the price includes all those migrations and refreshes, and it includes us undergoing 6 monthly ISO27001 audits and our own internal assessments, and it includes the escrow copy of the data, which the customer can take away with them should they decide to leave the service.

So the important thing is that the cost of long-term retention is predictable and cost-effective, it can be fixed, and it includes all the actions needed at the file level for proper preservation.

Conclusions

- Long term record retention and access needs digital preservation techniques
- EDRM systems don't necessarily have the required features
- Adopt a simple approach
 - Preservation standards and risk management
 - Storage, formats and metadata
 - Start from the bottom up and establish a burnline

Contact

Matthew Addis, Arkivum

matthew.addis@arkivum.com

Eldin Rammell, Rammell Consulting

eldin.rammell@rammell.com