Arkivum
Every bit archived

ArkivumLimited
R21 Langley Park Way
Chippenham
Wiltshire
SN15 1GE
UK

+44 1249 405060

info@arkivum.com
@Arkivum
arkivum.com

The Benefits of Archiving and Seven Questions You Should Always Ask

Whitepaper

Introduction

Long-term retention and access to data is a skilled business that shouldn't be undertaken lightly. The benefits of doing it properly can be substantial: meeting compliance requirements, saving costs, or monetising archive assets. But the cost of doing it wrong can be even higher. This whitepaper explores some of the benefits that come from a good archiving strategy and some key questions you should be asking to make sure whatever approach you take will allow you to realise these benefits rather than create a lot of sleepless nights!

Benefits of Archiving

There are many benefits to archiving data, especially when using the Arkivum service with its data safety guarantee. Just some of the benefits of a good archiving strategy include:

- Reduced cost

  This might be as simple as saving space on expensive storage used for production systems or it could come from in the longer-term, through use of historical data to better inform product development, service provision, customer management and many other business activities.

- Reduced risk

  Risks might be the inability to fulfil a compliance audit, the inability to retrieve data fast when the business needs it, the failure to meet retention or data protection requirements, or the risk of the costs incurred for recreating or replacing lost data (if that's even possible).

- Reuse of data

  Often data isn't just kept just for compliance - it can have value to the business. This might be monetising old film or video footage, use of in-service data to improve the manufacturing or maintenance of a product, or use of historical data for better CRM or targeted advertising.

- More manageable expenditure

  In-house archiving requires equipment, space, power, cooling and most importantly skilled people. Capital investments can be infrequent, but high. Operational expenditure models become an attractive alternative.

- Freeing up resources

  Long-term retention of data is typically not a core business activity for many companies that create or use data. So why divert valuable company resources to do this in-house? Outsourcing to a professional archive service provider like Arkivum means buying into not just dedicated infrastructure, but also a skilled team of data retention and access specialists.

Title
The Benefits of Archiving and Seven Questions You Should Always Ask

Part no
ARK/MKTG/ALL/179

Version
1.0

Date
Sept 2013

Status
Release

Seven Questions You Should Always Ask

Whilst there are many benefits to effective archiving, failures to select and manage the right solution can be very costly. Valuable assets can be lost, time and money is wasted, compliance audits can't be met, certification can be lost – it can all turn into a nightmare! Peace of mind comes from asking the right questions from the outset and making sure whatever solution or service you choose can deliver against these questions.

Q1. Can you be sure of the integrity of your data?

Are you guaranteed that every single bit of data you pull back from your archive is identical to what you put in - and can you prove it?

We generate checksums on all data we store as soon as it enters our systems. We check and agree these checksums with the customer so we are confident that we have correctly received the data (there's nothing worse than garbage in – garbage out!). We can digitally sign the list of checksums if needed for an auditable hand-over of custody. We then use checksums to manage data integrity by regularly checking checksums to detect any problems and then fixing them using the replicas we store for each and every file. And we make all this information available to the customer so they have a full record of when their data was last checked and that all the copies we keep are in tip-top condition. Whenever a customer asks for a file from the archive we will use the checksums to confirm that it is bit-identical to the original – no files are ever returned without passing this test!

Q2. Is your data always confidential?

Can you be sure that there is no way for anyone else to access and understand your data - and that includes any service providers used to hold the data?

We have a full information security management process in place that covers all aspects of security including people, process, facilities, equipment and software. Recognising that security comes from defence in depth using layers of protection, just some of the measures we use include are: all customer data is encrypted on-site with AES encryption before it leaves the customer network. The customer owns the key-pair and Arkivum does not have a copy of the keys at our data centres or offices, so we have no way of decrypting customer data. Data is then transferred to our data centres over a two-way authenticated VPN connection. Data storage at our data centres and escrow site uses a range of physical security measures. Remote online access to this data is only available to our customers and hardware firewalls are used to tightly control both customer access and Arkivum's management of our service. Each customer's data is on a dedicated set of tapes, including in escrow. Rest assured no one is able to read your data except you.

Q3. Do you know the authenticity of your data?

Do you know who first put the data into the archive, who's then accessed that data, and who has handled the data inside the archive to keep it safe and secure?

We store a complete record of where every file in our service has come from, what its ownership and permissions are, when it has been retrieved by the customer, when we've migrated it to new media, when we've checked its integrity, when it went into escrow – as well as any unauthorised attempts to access the file. For information that is customer sensitive, e.g. file permissions, name or path, then this information is encrypted and signed with the customer's key before it leaves the customer network. This allow assertions to be made that this information is secured and hasn't changed, whilst allowing the information to also be stored offsite in escrow along with the data itself. In that way, the escrow store includes all information about the customers data without compromising confidentiality or allowing undetected changes to be made that would jeopardise authenticity. Records of customer data access or any activities within our service such as integrity checks that 'touch' the data are all held in secured log files that themselves are replicated and signed by Arkivum to ensure that a full audit trail is available for all customer data.

Q4. Is your data available when you need it?

If you need data quickly then can it be retrieved in a matter of minutes? Can this be done for every single file you want to store?

Files in our service are stored in two separate data centres in online tape libraries. If there is a problem within one of the data centres or the network access to it, then the Arkivum appliance that sits on the customer's network will automatically fail-over to the other data centre for both ingest and access. When the problem is resolved, then the files in the two data centres are automatically synchronised –and all along from the customer point of view it looks like nothing has happened. The use of tape libraries mean that access to files is quick – less than 5 minutes to get to the start of any file and then very high data rates when getting data back to the customer's appliance. It will often be that the network connection is the bottleneck – not the speed of our service. Sometimes for very large files only a small part of the data might need to be retrieved by a customer. This is sometimes called 'partial restore'. We can retrieve just the data that's needed, which avoids having to move files that are GBs or even TBs over the network just so a small part can be used at the customer site. All this combined means that we offer an SLA not just of 100% data integrity, but availability of >99.9% per month and file access time of less than 5mins. If files are accessed frequently, then they can be cached on the customer appliance to provide fast repeated access without needing to pull data over the network every time.

Q5. Do you know who is in possession of your data at all times?

Do you know exactly where your data is stored and who runs the facilities? Do you know who controls how your data is stored and who is responsible for its safekeeping?

All customer data is stored in fixed locations that are known to our customers. We never move customer data to another location without permission first. Data for each customer is on dedicated tapes and customers can even go into one of our data centre to physically see these tapes for reassurance if they need to. These tapes are only ever handled by trained Arkivum staff who are responsible for their safekeeping through use of our highly automated data management systems. We use a barcode based tape tracking system so we know the exact location of every tape we hold. This is done all the way from blank media coming into our facilities for QA through to loading tapes in our Data Centre tape libraries and then any secure transfer of externalised tapes to our escrow site. We record exactly who at Arkivum carries out each activity and all our staff adhere to tightly defined company policies that embody digital preservation and information security recommended practice.

Q6. Are you confident that you can always use your data?

What happens if your vendor or service provider goes out of business or stops support? Are you locked into proprietary data formats or tools? What happens if you want to change applications that use your archive - do you need to go through a complex systems integration process?

Our use of data escrow gives complete peace of mind that your data is accessible to you no matter what happens to our business or service. All data is stored on data tapes at an independent third-party escrow site. We use LTFS so access to the data on each tape is simple and can be done using either a library or standalone drive available from a wide range of vendors. Each tape comes complete with a full manifest and set of open source tools that can be used for easy recovery of the data. The manifest contains a complete index of what files are on the tape, the permissions, paths, checksums and encryption used. The manifest is human readable and supplied on separate media so it can be accessed and inspected without having to physically load up a tape. The open-source and cross-platform tools can also read the manifest and automate the process of retrieving the files from the tape and decrypting them (using keys provided by the customer at this point).

Q7. Can you always assert the provenance of your data?

Faced with a compliance audit, the need to understand whether you're litigation ready, or the need to prove the authenticity of your data, can you provide an audit trail that says what the data is, where it's come from, and everyone and everything that's happened to it since?

We have a comprehensive approach to recording and agreeing data integrity checksums, recording and providing a complete record of how we have handled all customer data, tracking each and every customer access to data, and logging any attempts at unauthorised access – either physically

or over the network. We know exactly what every data file is, its origins and who/where/when any preservation actions have been applied. Together with our information security management system and tightly defined policies, we can provide a remarkably complete and rigorous record of everything and everyone who has interacted with your data in any way, be it access you have made or actions we have taken to ensure integrity and authenticity.

Summary

These aren't just questions that we've just made up in order to highlight how good our service really is. The questions about confidentiality, integrity and availability are the bedrock of information security and are known as the CIA triad. The questions about authenticity, possession and utility are further essential facets of risk management for information assets. The final question (provenance) is a pillar of digital preservation, but is probably more familiar in a business context as being able to support auditability — it is about asserting a chain of custody for data assets.

If you want to know more about these areas, then have a look at ISO 27001: Information Security Management System[1] and the new ISO standard 16363 for Trusted Digital Repositories[2]. It almost goes without saying that with Arkivum's assured archiving service you can confidently answer 'Yes' to every one of the seven questions above.

[1] http://en.wikipedia.org/wiki/ISO/IEC_27001
[2] http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=56510