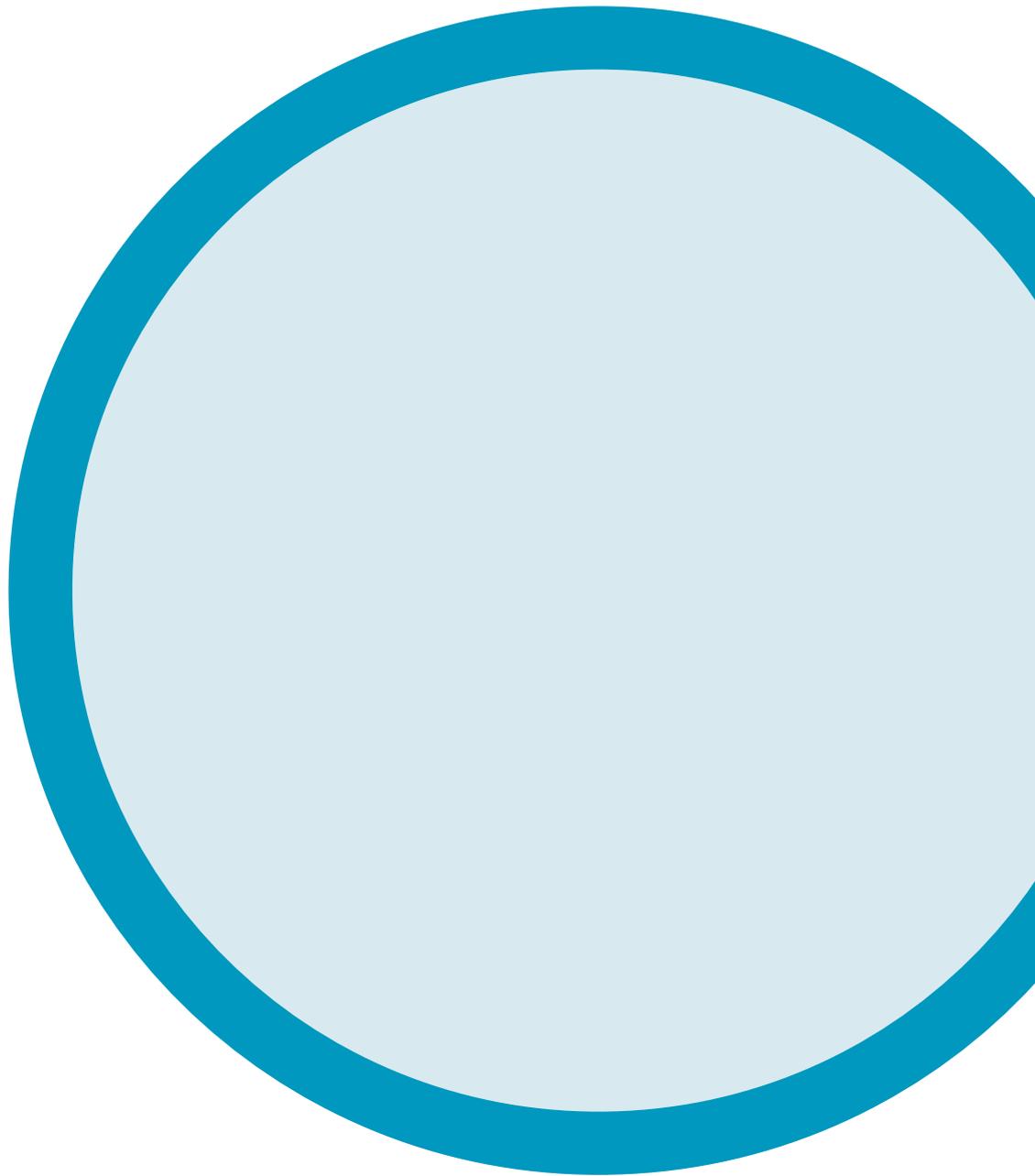


An Emerging Market:

Establishing Demand for
Digital Preservation Tools and Services



Executive Summary

Digital data are all pervasive. Across industry, government and the public sector, digital files are vital to modern business. They facilitate communications, allow efficient exchange of information and reduce running costs. Digital data appear to be just a mouse-click away, whenever we need them, but for how long?

Like analogue records, digital data are susceptible to decay. Bits and bytes may disappear over time and hardware and software can become obsolete, leaving existing data unreadable. Storage deterioration and lack of long-term management undermine the readability and accessibility of digital content. Data loss may have serious repercussions for an organisation. It could mean the loss of valuable research data, removing an organisation's competitive advantage. Or it could result in the disappearance of crucial audit trails with undesirable financial and legal consequences for any regulated organisation. It is therefore vital to carry out active digital preservation. That is "to ensure ongoing, meaningful access to digital information for as long as it is required"¹. This is not just a protective measure, it is also a business opportunity: a new market for digital preservation solutions and products.

This Planets white paper examines digital preservation issues (strategic, technical and economic) from the perspective of vendors and suppliers. It draws on qualitative analysis of 18 interviews with leading IT companies based in the US, Europe, the Middle East and Australasia. Their thoughts and opinions are summarised in this paper. They shed light on the new emerging digital preservation market.

A Planets White Paper by
Pauline Sinclair, Tessella and Amir
Bernstein, Swiss Federal Archives

Published July 2010

Key Findings

- The digital preservation market is still in its infancy; there is plenty of potential for growth as it affects all business sectors.
- Currently, engagement is led by the memory institutions² at national and international level. Engagement is also high in government and research organisations and emerging in the private sector.
- Legal obligation is the key driver for organisations to engage in digital preservation, although additional motivations vary by sector.
- Although digital preservation is business critical, many organisations do not have a policy to cover it. Where policies exist, their comprehensiveness is variable.
- There is a lack of information on the costs of digital preservation and its benefits (both tangible and intangible) which makes it hard to put together a convincing business case.
- Budgets for digital preservation are generally short-term and tend to be project-based.
- There is a perceived immediate need to preserve documents, images, audio, websites, video, spreadsheets and emails.
- Migration is strongly preferred to emulation to preserve digital material. However, this may change as emulation has a distinct role to play and there is some interest in emulation tools amongst the briefings' participants.
- Participants thought that the most important factors for a digital preservation solution are that it should: maintain authenticity, reliability and integrity, adhere to metadata standards, and check records have not been damaged.
- Scalability of digital preservation solutions to high volumes of content and high ingest rates are regarded as important but scalability of access is not yet important.
- While attendees thought standards are important in digital preservation, particularly OAIS and ISAD(G), they also thought that there are currently too many standards.
- There is still a need for guidance (particularly training), exchange of best practice, and awareness-raising through conferences and workshops.
- There is confusion as to what is meant by digital preservation. The difference between passive (storing multiple backups of data) and active (using migration or emulation to provide access to obsolete formats) preservation needs to be clarified. It should be emphasised that only active preservation can ensure that digital material can be accessed in the future.

¹ "Mind the Gap. Assessing digital preservation needs in the UK", 2006, Waller, M., and Sharpe, R., Digital Preservation Coalition.

² 'Memory institutions' is a collective term for museums, archives and libraries.

Methodology

In 2009, Planets conducted a series of interviews with 18 IT companies, who between them cover all aspects of the digital preservation market, in order to better understand that market. In-depth structured interviews were held with each of the participants. Two-thirds of the interviews were held in Brussels and the remainder were conducted over the phone; all were held over a period of 3 weeks. Each interview covered exactly the same topics with the facilitators following a script to ensure there was no bias. Initially, the participants were asked quantitative questions, where their choice of answers was constrained, and then these answers were explored further in open discussions. The questions were grouped into three sections covering: the digital preservation products and services offered by the participants, their views on the digital preservation needs of the market, and opportunities for Planets and the participants to work together.

Of necessity, only a relatively small number of suppliers and vendors could be interviewed. However, the participants were picked to form a representative sample of the IT market in terms of size, market coverage and involvement in digital preservation. So, while the quantitative results reported in this white paper are not statistically significant, they are indicative of the trends and so offer a useful insight into the digital preservation market.

The Participants

The 18 participants in the Planets Supplier and Vendor Briefings comprised senior representatives (including 6 CEOs and 8 IT managers) from IT companies across the digital preservation market. The participating companies, who between them represent millions of customers, are mainly based in Europe but also in North America, the Middle East and Australasia. They range from small companies with up to 75 employees, an annual turnover of 0.5 to 4.5 million € equivalent and activities on a national or a regional level (two or three countries) to multinational giants with thousands of employees and an annual turnover the equivalent of billions of euros.

The companies derive between 10% and 80% of their revenue from digital preservation services and have been involved in digital preservation from 2 years up to 25 years. They all supply clients in the archives, libraries, public administration and commercial sectors and cover a broad spectrum of content segments, ranging from the sciences and medicine, via economics and the law to the arts and humanities.

Digital preservation suppliers are starting to follow the trend of providing services on-line and storage in the cloud, with 9 of the 14 participants who answered the question currently doing so. Of these, two-thirds offer data storage or preservation-planning services, only a third offer data analysis and one or two offer data normalisation, data migration or emulation services. The majority of the participants receive data on-line and subject it to a range of quality checks before accepting it.

The Market

What is Digital Preservation?

Digital preservation means different things to different people. The main source of confusion is between the passive and active approaches to preservation. The first usually refers to the long-term storage of data. This approach aims to preserve the bit-streams through multiple backups. It ensures data are retained but, as systems are regularly replaced and formats constantly change, this approach cannot guarantee that the data will be readable or understandable in the future. Active preservation starts with the passive approach but uses either regular data migration or the provision of emulation tools to ensure that data can be read, accessed and understood for as long as necessary. For many this is seen as too complicated and so it is the passive approach that wins: "Most institutions regard preservation as storage."

In the US, digital preservation tends to mean passive preservation, as exemplified by LOCKSS (Lots of Copies Keeps Stuff Safe), while in Europe the active approach is more common and this is the approach that Planets has taken. Participants also identified differences in attitudes by sector. In the private sector, the emphasis is on storage and immediate access. In the public sector, and particularly memory institutions, it is on active preservation and long-term access. There is also some confusion between preservation and digitisation. While digitisation makes material visible and disseminates it widely, it does not in itself guarantee long-term availability.

«There is still a lot of confusion – a lot of work in education to be done.»

Source: 2009 Planets supplier / vendor briefing participant

A Market in its Infancy

The market is perceived to be at a relatively early stage in its development. Engagement in digital preservation is being led by the memory institutions and is also thought

to be high among government and research organisations. This is reflected in the customer sectors in which participating companies have operations.

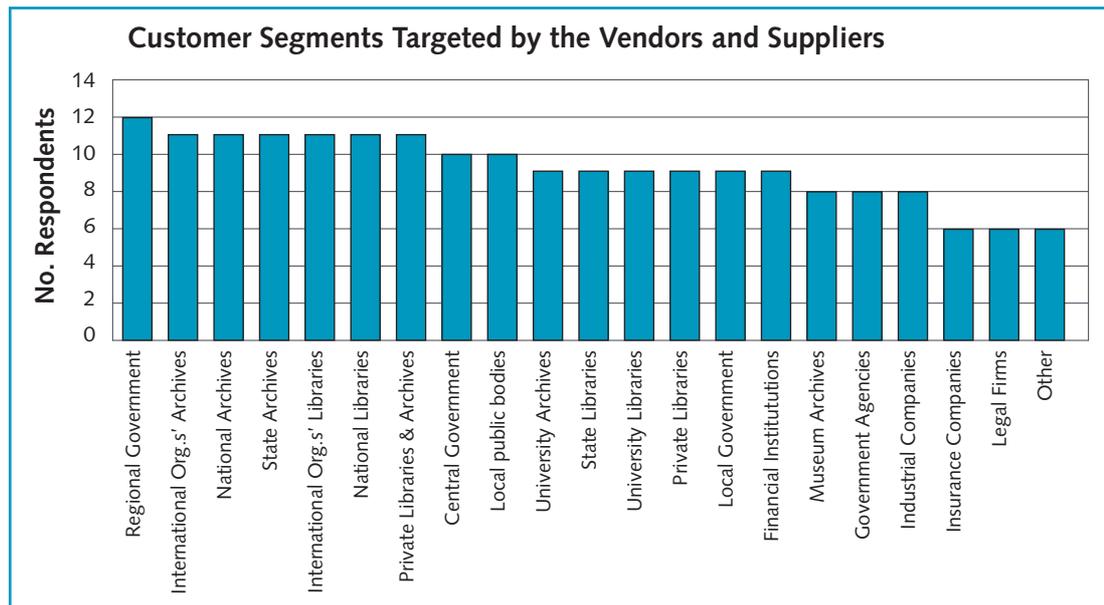


Figure 1: Customer Segments Targeted by the Planets Supplier and Vendor Briefings' Participants

There is emerging engagement in digital preservation in other market sectors, specifically:

- The financial sector, with banks and insurance companies obliged to retain audit trails and which, since the 1980s, have been moving towards digital record management systems.
- Healthcare organisations, which are required to retain patients' records.
- The pharmaceutical industry, which is interested in preserving accumulated scientific knowledge.
- The oil industry, which is looking for possibilities to reanalyse old seismographic data in search of new oil and gas reserves.
- The photographic industry, where the growth in digital images has been and continues to be huge.

«10% do digital preservation, 70% are thinking about it but don't know what to do and 20% are not aware of it.»

Source: 2009 Planets supplier / vendor briefing participant

Engagement is also perceived to be led by large organisations with capital and in-house or third-party IT resources. This is unsurprising as the digital preservation market is still in its infancy and so economies of scale have not yet been brought to bear; it is only those with plenty of resources and strong motives who have started to tackle it. Engagement is thought, by some, to be higher in the US than Europe.

Legislation introduced over the past decade has required organisations to retain digital information for significant periods of time (ranging from 6 months to 10 years or more). Such legislation includes the Data Protection Directive 95/46/EC, as enacted in the countries of the EU, and MiFID in Europe, Sarbanes-Oxley and HIPAA in the US and the Basel II Accord worldwide. It applies not just to organisations' core business data but day-to-day administration data such as contracts, pension plans, health and safety records etc. Financial institutions are particularly concerned with compliance and the need to keep audit trails of transactions to minimise liability.

While legal obligation is the key driver for most organisations, data security, business continuity and re-analysis of data are also considered to be important motivations for engaging in digital preservation. Different drivers are perceived to be at work in different sectors; a secondary driver for many in the private sector is the commercial need to reanalyse data. Outside memory institutions, engagement is motivated by external imperatives rather than a belief in the inherent value in retaining information.

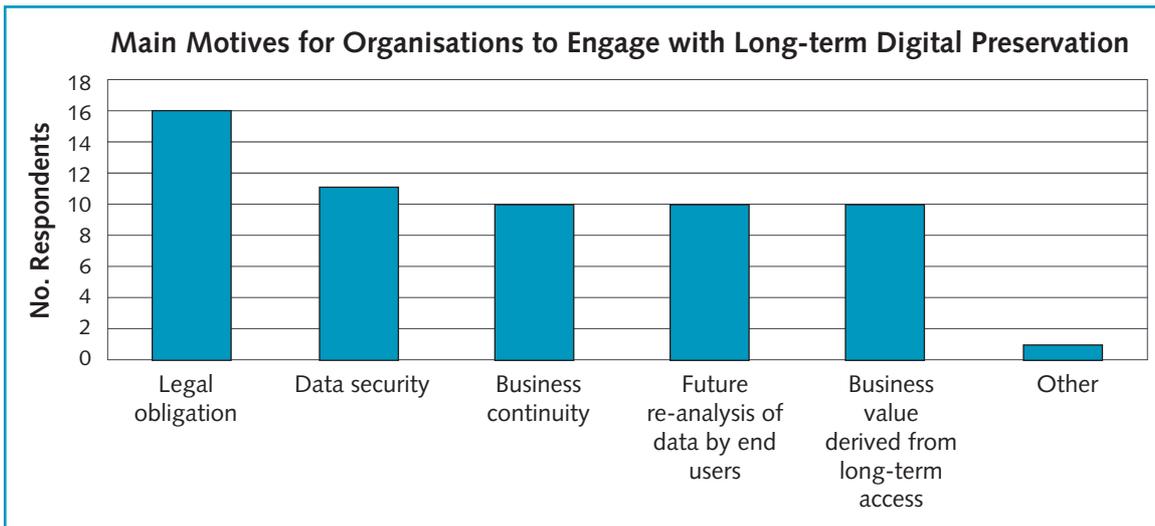


Figure 2: Main Motives for Organisations to Engage in Long-term Digital Preservation

The Planets Market Survey³ demonstrated the correlation between articulating a digital preservation policy and engagement in digital preservation activities. However, the briefings' participants saw things as less clear-cut. Although some of their customers have such policies, many have only inadequate policies or none at all. For some, this is because they are dependent on a single vendor's solutions and so don't see the need to write their own policy, while others are just getting on with it. The public sector and memory institutions lead the way with the development of policies and are perceived to have policies that are comprehensive and that relate to long-term preservation. Where policies exist in the private sector, they may be incomplete or concerned with storage and access only. As with engagement, it is often larger companies with their own IT resources that have developed a policy, while small companies lack one. Organisations in the US are less likely to have a policy than organisations in Europe, despite greater pressure to retain data.

Articulating a digital preservation policy is only the first step, though, and as one representative noted: "Writing policies is cheap, implementing them is expensive". Unfortunately, digital preservation is often seen as a luxury, particularly in the current economic climate. The case has not been made to prioritise digital preservation and consequently it is subject to economic conditions. There is a lack of data on the costs of digital preservation, both initial set-up costs and long-term running costs, and the benefits (such as data reuse, avoidance of fines from regulators, compliance with legislation), both tangible and intangible, have not been clearly articulated. All of which makes it hard to put together a solid business case.

It is not surprising, therefore, that where digital preservation budgets exist, they are short-term: typically 1 to 3 years ahead and rarely more than 10 years. Although memory institutions and Government are thought to take a longer-term perspective. This is a problem when legislation demands that organisations

retain data for longer periods than the budgets cover, as it hinders planning. In many cases funding is strictly project-based. This may be no bad thing, as it can lead onto more long-term funding when a case can be made to sustain the digital preservation system created in an initial project.

«Digital preservation is still considered a one time project – usually capital budget and not revenue budget is used.»

Source: 2009 Planets supplier / vendor briefing participant

In some cases, budgets may be spent preferentially on digitisation. Several participants commented that end-users are more interested in ensuring wider access to, and greater visibility of, their existing material through digitisation than in engaging in digital preservation. "Most want to put their material online and are not interested in preservation." Through such actions end-users believe they have preserved their analogue data but, in the long term, the digital surrogates created by digitisation need to be preserved, just like born-digital material. One vendor commented that the focus on digitisation is the result of funding being available to support it: "Funds [exist] for digitisation but not for preservation. This may be the wrong order but, on the other hand, it is pragmatic. They digitise and then look for strategies and solutions to retain the content."

³ "The Digital Divide: Assessing Organisations' Preparations for Digital Preservation", A Planets White Paper, 2010, Sinclair, P.

The Nature of the Solution

The briefings' participants thought a wide range of factors were important for end-users when evaluating a digital preservation solution. Factors considered important or very important by more than three-quarters of the participating suppliers and vendors were: maintains authenticity, reliability and integrity, adheres to metadata standards, checks records have not been damaged, retrieves content by description, is able to store many different types of content, and characterises records by extracting metadata.

Whether or not networked services are important depends on the end-user. Small companies, whose core business is not archiving, are not looking for software or a solution to install in-house but an on-line service to handle the small quantities of data they need to preserve. For them, networked services are very important. However, for organisations with large quantities of data to archive, of the order of terabytes or petabytes, networked services are useless as the data volumes would overload the network. Additionally, not all organisations trust someone else to look after their data, particularly if it is confidential or commercially sensitive.

Scalability of digital preservation solutions to high volumes of content and high ingest rates are important, while scalability to high access rates is not thought to be so important. This may reflect the embryonic nature of the market, with end-users currently being concerned with gathering and storing the digital information rather than disseminating it widely. However, for many organisations the data they need to store is confidential and has a limited audience, so high access rates will never be important.

The attendees agreed that standards are important, particularly OAIS⁴ & ISAD(G)⁵ but several commented that currently there are too many standards. This plethora of standards has several sources. There are the archiving metadata standards, such as Dublin Core, PREMIS and METS, the electronic records management standards, such as ISO15489, MoReq2 and DoD 5015.2, and other relevant standards such as those for digital archiving (e.g. OAI-PMH, OAI-ORE, TRAC), IT security (e.g. ISO17799, ISO27001, ISO72002), and domain-specific standards (e.g. BIP0008 for UK courts of law).

«Standards are on the whole highly important. Unfortunately there are too many standards, which threaten to make standards themselves redundant.»

Source: 2009 Planets supplier / vendor briefing participant

As for the strategy used for active preservation of digital files, migration was overwhelmingly preferred to emulation (over 85% of participants said their end-users only use migration). This is due to the perception that migration is more sustainable, less resource hungry, and easier to control since it is easier to see what has been done and compare the results. By contrast, emulation is seen as being too complicated for many types of data, difficult to manage, and there is only limited knowledge of how to deal with it.

Despite the reservations about emulation, eight attendees expressed interest in Planets emulation tools to ensure access to old records for decades to come. Attendees recognised migration and emulation have distinct roles and both strategies have a place in active digital preservation. Potentially this could break the vicious cycle surrounding emulation, whereby the lack of familiarity with emulation and its perception of being hard to implement, leads to a lack of demand for emulation tools which, in turn, means a lack of products and effort to educate end-users. This is important as both techniques address different digital preservation needs. Emulation is particularly important for providing access to digital objects which have dynamic behaviour or where users need to interact with the object, such as multimedia materials, geographic information systems (GIS) and educational software.

Whatever digital preservation solution is implemented, it will need to preserve a broad range of digital material. The participants perceived an immediate need to preserve documents, images, audio, websites, video, spreadsheets and emails, with different types of business prioritising different object types. Emails are an interesting case as they are treated legally as correspondence and thus covered by recent legislation. So organisations are required to retain them for specific periods of time. However, their preservation is complicated by the need to handle the wide range of attachments that can be associated with them. As one vendor said: "Email preservation is very interesting but seldom dealt with." Websites are also seen as being important but difficult to preserve. There are different approaches to preserving them, with one vendor recommending preserving the database behind the website and another taking snapshots. Given the broad range of material needing preservation, it is perhaps not surprising that participants saw potentially high interest in a service to recover obsolete files, a service that is currently lacking.

⁴ 2003 Reference Model for the Open Archival Information System

⁵ General International Standard for Archival Description

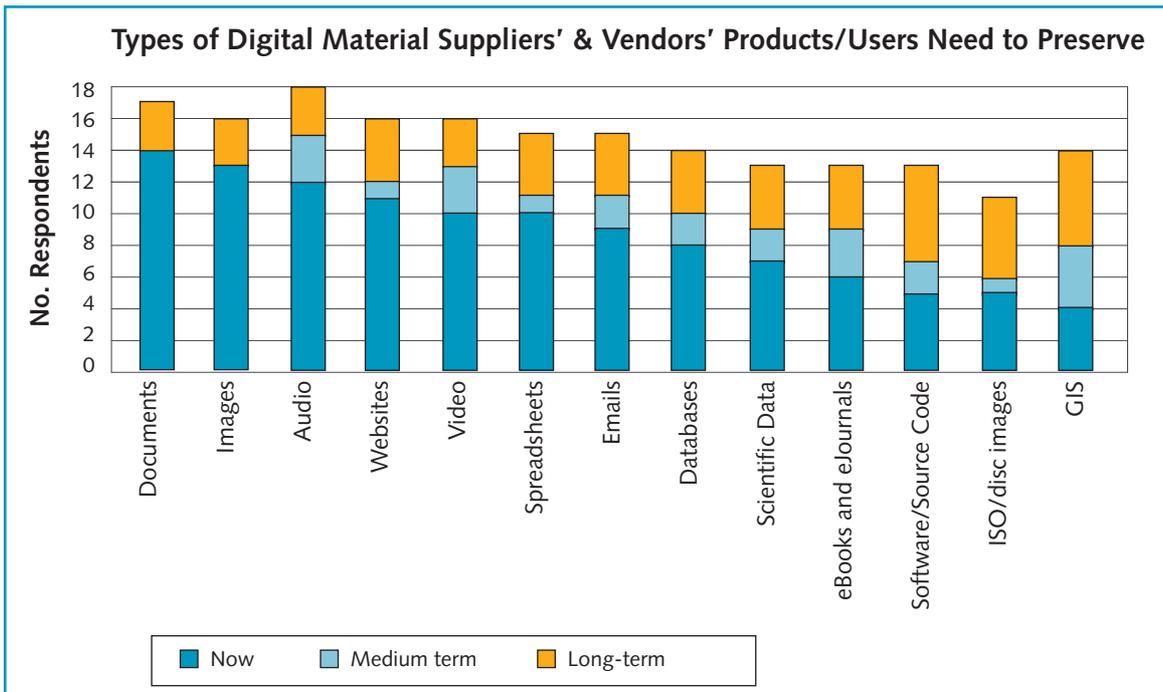


Figure 3: The Types of Digital Material that Vendors' and Suppliers' Products/Users Need to Preserve

Opportunities

Participants recognised that there is an emerging market in digital preservation with the potential for growth as the need to preserve digital content affects all business sectors. Most said that they had an active interest in developing the digital preservation market or tapping into existing demand through provision of tools and services. They recognise that there is scope to stimulate demand through development of archiving standards, provision of strategic advice on implementation, and development of tools for planning and executing digital preservation schemes.

Participants in the briefings were clear that there is still considerable work to be done to raise awareness about the importance of preserving digital information for the long-term. This needs to start by clarifying exactly what is meant by digital preservation: both the difference between active and passive preservation and that digitisation is not digital preservation. In particular, the consequences of not carrying out digital preservation need to be emphasised.

Outside memory institutions awareness falls and management, in particular, needs to be made more aware of its importance. As one consultant put it: "[Digital preservation] is business critical, but it's not being thought about strategically. There is a need for a business case – scare stories – to be articulated and to motivate organisations to deal with digital preservation." For those aware of the importance of digital preservation there is a need for guidance, the exchange of best practice, and help with creating a business case. Participants see a demand for

conferences, workshops and user groups, due to their role in sharing information. Practical measures include: providing access to preservation planning tools and allowing users to test planned preservation actions in a controlled environment on sample data.

Attendees drew attention to areas where they would like to see change. In particular, they would like to see cost models for digital preservation which are currently missing and are necessary for building a solid business case. Such models need to put a price on both the costs (e.g. loss of audit trails) and the benefits (e.g. re-use of data), and also demonstrate the price of inaction.

In addition, participants think the current proliferation of standards needs to be curbed and some consolidation take place. Of course this will require agreement on which standards are important and need to be kept, which are specialised and should no longer be universal, and which are no longer needed and should be eliminated.

The briefings highlight a number of opportunities that are ripe for exploitation. The most immediate need is for digital preservation consultancy, providing guidance on what needs to be archived, how to do so and how to build a business case for digital preservation. More digital preservation tools for characterisation and migration are required to cover the vast range of formats that digital material comes in. There is potentially a niche market providing on-line digital preservation services to smaller companies for whom archiving is not a core activity. Finally, there is a need to educate the digital preservation community about the role emulation has to play which needs to be backed up by the provision of practical emulation tools.

Summary

The digital preservation market is still in its infancy. However, the relevance of digital preservation is extending far beyond memory institutions into all business sectors. For many, though, digital preservation is seen as an onerous obligation not an opportunity to realise the long-term value of existing resources. Memory institutions are culturally predisposed to retaining information and see inherent long-term value in it. By contrast, other sectors are having this imposed on them from outside, by legislation or business pressures, and regard it as 'another drain on the bottom line'. This is manifested in the lack of digital preservation policies, or their incomplete nature, and the short-term project-based nature of budgets, which vanish in times of economic uncertainty. It is also seen in the confusion about what digital preservation entails.

Solutions already need to deal with a wide variety of content and this will only increase in the future. It is therefore no surprise that scalability to high rates of ingest and volumes of content are important now. Although preservation is not yet deemed to be necessary for as long as 50 years, the ability to trust a preservation solution to keep an object intact and accessible is regarded as important. The need for preservation standards is recognised, as is the need for their rationalisation. While there is a strong preference for migration, there is emerging recognition that emulation has a place for particular types of content.

Future engagement requires work to be done to raise awareness about the importance of preserving digital information and to articulate a business case for it. At the most basic level there needs to be a clear definition of what digital preservation is. Following on from that, the costs and benefits of digital preservation need to be set out and the cost of no action clearly demonstrated. Further guidance on how to implement digital preservation is required and there is a demand for information and training. There is an ongoing need for the development of tools and services, in particular to cater for the needs of smaller organisations with fewer resources. Finally, workable approaches to emulation need to be developed and so predispose end-users to use it where appropriate.

Acknowledgements

Planets would like to thank the individuals and organisations who attended the supplier/vendor briefings and also the British Library and the Swiss Federal Archives for undertaking the work on behalf of the project.

About Planets

Planets (Preservation and Long-term Access through Networked Services) was a four-year project co-funded by the European Commission under the Information Society Technologies (IST) priority of the Sixth Framework Programme for research, technological development and demonstration (IST-033789). Between 2006 and 2010, Planets has worked to develop a framework and a suite of practical tools and services that will enable institutions in Europe to manage and access their digital collections for the long-term. Co-ordinated by the British Library, the project has brought together the expertise of 16 national libraries, archives, research institutions and technology companies in Europe.

Planets technology provides access, through a single open-source application, to a range of Planets and third-party digital preservation tools and services which support and automate a range of processes. These include creating a preservation policy, planning to preserve specific content or collections, identifying the significant properties of collections and individual objects to be preserved, assisting with the identification and selection of tools and services, validating action that has been taken and determining the extent to which it has been successful.

In addition, Planets has developed a controlled experimental environment where tools, services and workflows can be tested using pre-defined sample content to assess their suitability for use in digital preservation. The outcomes of such experiments will be used to update information available to Planets users about the tools' appropriateness to preserve particular content.

Planets' results will be maintained and developed by the Open Planets Foundation (OPF), an independent, not-for-profit company. OPF's members share a commitment to ensuring long-term access to digital content. OPF will provide hosted access to Planets services, technical support and training. It will also coordinate development of Planets services, tools and technology by supporting and engaging with the Planets Open Source community. Members benefit by sharing experience and know-how in a community of experts.

Further Information

For more information about Planets visit:
www.planets-project.eu

To join our user community and receive regular updates about Planets, register at:
www.planets-project.eu/community

To download Planets reports and newsletters visit:
www.planets-project.eu/publications.

To find out more about the Open Planets Foundation, including membership, visit:
www.openplanetsfoundation.org.

You can email your questions to us at:
info@planets-project.eu