The National Archives

The digital landscape in government 2014-15

Business intelligence review

Published: February 2016

OGL

1. Contents

2. Glossary

BIS	Department for Business, Innovation and Skills
CTO	Chief Technology Officer
CO	Cabinet Office
CCS	Crown Commercial Service
CTS	Common Technology Services
DCMS	Department for Culture, Media and Sport
DH	Department of Health
DPA	Data Protection Act
DRI	Digital Records Infrastructure
DROID	Digital Record Object Identifier
DTP	Digital Transfer Project
EDRMS	Electronic Document and Records Management System
ERM	Electronic Records Management System
FOI	Freedom of Information
GB	Gigabyte
GDS	Government Digital Service
HMRC	Her Majesty's Revenue and Customs
IAR	Information Asset Register
IM	Information Management
IMA	Information Management Assessment
KIM	Knowledge and Information Management
MB	Megabyte
MoJ	Ministry of Justice
OSP	Operational Selection Policies
PRA	Public Record Act 1958
RTR	Records Transfer Report
TB	Terabyte
WTK	What To Keep

3. Executive summary

The National Archives is one of the world's most valuable resources for research. As the official archive and publisher for the UK government, and England and Wales, we are the guardians of some of the UK's most iconic national documents, dating back over 1,000 years.

Our role is to collect and secure the future of the government record, to preserve it for generations to come, and to make it as accessible and available as possible. As well as bringing together the skills and specialisms needed to conserve some of the oldest historic paper documents, The National Archives is leading digital archive practices to manage and preserve government information in the past, present and future.

Over the last few years, The National Archives has been developing innovative new digital technology and practices to prepare for the transition from paper to born-digital records from government departments.

This report sets out the challenges in the current digital landscape including the volume and limited structure of born-digital records, technology for digital sensitivity review and the resources needed in government departments. It also sets out recommendations for government departments, such as developing strategies to move born-digital records and metadata between systems and identifying the minimum skills needed to provide a smooth transition into the born-digital archive of the future.

4. Context and background

The National Archives in the United Kingdom is a non-ministerial department sponsored by the Department for Culture, Media & Sport (DCMS). It is the official archive of the UK government and for England and Wales. Under The Public Records Act 1958 (PRA), public bodies are required to select records for permanent preservation, under the guidance and supervision of the Keeper of Public Records, and transfer these records to The National Archives (or an approved place of deposit) no later than 20 years after their creation. Under the Act, records are defined to include not only written material but 'records conveying information by any other means whatsoever'[1].

From 2016, public bodies will begin to transfer born-digital records – those records created originally in born-digital formats such as emails, documents and spreadsheets. Born-digital information should not be confused with digitised information, which consists of analogue material (e.g. paper, film, photographs) that has been rendered in digital form. In this review, when we speak about digital records or information we will be referring to born-digital information as opposed to digitised information.

Preserving digital records is a major challenge for archives across the world. The National Archives is leading the archive sector in embracing the challenges of storing digital information for future generations and providing access to this information to the public.

Building on years of research and experience in the field of digital records – in the areas of digital continuity, transfer, digitisation and preservation, big data, data visualisation and data linking – The National Archives is responding to the digital challenges to ensure that we are ready to keep the nation's public records safe and accessible for the future, whatever their format.

The chart below shows the cumulative number of government departments due to transfer born-digital records to The National Archives for the first time. This does not represent a cumulative yearly transfer amount – in megabytes (MB) or gigabytes (GB) – but rather the cumulative number of government departments due to transfer by 2021. The data collected indicates that, by 2016, there are expected to be 12 departments scheduled to transfer digital records, with an approximate total of 50 in 2021.

---

[1] The Public Record Act 1958: legislation.gov.uk/ukpga/Eliz2/6-7/51/contents

Figure 1.1 Cumulative number of government departments and agencies expected to transfer born-digital records to The National Archives for the first time under the Public Records Act (Source: Digital Questionnaire, April 2015)
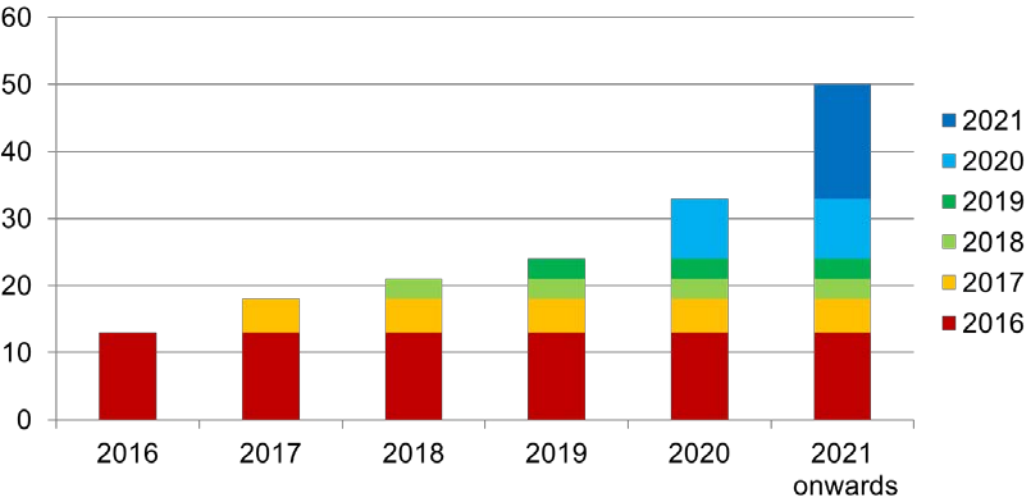
To ensure that The National Archives and government departments are ready for the first, large-scale, business-as-usual, born-digital record transfers, the Digital Transfer Project (DTP) was launched in September 2013. The aim of the project was to develop a scalable process for the transfer, ingest and presentation of born-digital records with long-term value; and to enable them to be held securely while closed and be accessible to the public when open.

As part of the DTP, a business intelligence review was undertaken to understand the current digital landscape and digital challenges of government departments in the UK. In particular, we wanted to understand

- the volumes and types of digital information being created and managed by government
- the capabilities and capacity of government departments to manage digital records, both current and legacy
- the ability of government departments to prepare for the transfer of digital records to The National Archives.

This information will help The National Archives and government departments to plan resources and prepare for the future transfers of born-digital records. The findings will also assist in determining areas where government departments may need support during the transfer process, enabling potential challenges to be identified and addressed at an early stage.

4.1   How was the business intelligence review conducted?

The information in this report has been compiled from:

- a digital questionnaire from April 2015 with replies from 60 departments, including 19 of the key departments[2]

- digital questions collated as part of the October 2012 Records Transfer Report (RTR)[3]

- a digital questionnaire from November 2012 to January 2013 with replies from 77 departments, including 12 of the key departments

- Information Management Assessment (IMA) reports from 2012-13

- IMA lessons learned report March 2015[4]

- IMA good practice report September 2014[5]

- 19 of the key government departments interviewed between July 2014 and April 2015.

Where this report refers to the key 21 departments or organisations it is referring to those that The National Archives have identified as accounting for 90% of the records transferred to The National Archives over the last three years.

The research and writing of this report was undertaken by an expert team at The National Archives experienced in digital records management. This research was undertaken prior to the publication of Sir Alex Allan's review of government digital records and seeks to complement the review's work.[6]

4.2   Structure of the report

This report is structured according to the Gateways (or stages) of the digital transfer process:
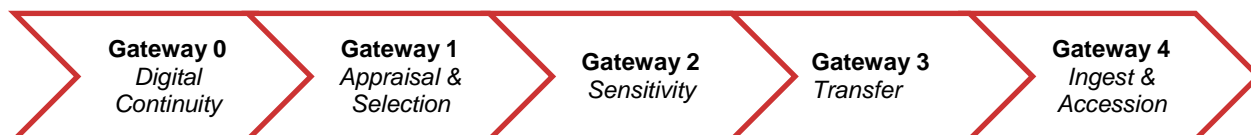
Gateways 1 to 4 were originally created to ensure that paper records were being transferred according to the most efficient process. The principles behind 1 to 4 are conceptually the same for both the paper transfer process and the digital transfer process. Gateway 0 was developed and included to guide departments in the care and preservation of their digital records until they are transferred to The National Archives.

These gateways are explained in greater detail later in this report along with the areas relating to each Gateway that were investigated during the review. The business intelligence review did not collect any information or data relating to Gateway 4 (Ingest and Accession), as this gateway mainly relates to internal processes within The National Archives.

---

[2] These 19 departments were taken from 21 organisations that The National Archives have identified as accounting for 90% of the records transfers to The National Archives over the last three years

[3] nationalarchives.gov.uk/about/our-role/transparency/record-transfer-report/

[4] nationalarchives.gov.uk/documents/information-management/ima-programme-lessons-learned.pdf

[5] nationalarchives.gov.uk/documents/information-management/ima-programme-good-practice-report.pdf

[6] Independent Review of Government Digital Records Sir Alex Allan

5. Digital continuity (Gateway 0)

Digital Continuity[7] is the ability to use digital information in the way that you need for as long as you need to. Digital information is useable if you can

- find it when you need it

- open it as you need it

- work with it in the way you need to

- understand what it is and what it is about

- trust that it is what it says it is.

Digital continuity is about ensuring the ongoing access of digital records which have long-term value at government departments, until they can be transferred to The National Archives.

The creation and management of digital records can impact their availability and the ability of users to understand the content and context of the records.

5.1   Departmental responsibilities relating to Digital Continuity

Government departments must ensure that the digital public records they create remain accessible, readable and useable until transfer to The National Archives. Departments should know what digital records they hold and monitor the state of their digital records. This will enable them to identify any risks that might affect the ongoing availability of their digital assets.

5.2   Areas the business intelligence review examined related to Digital Continuity

- Technologies in use, digital records formats and volumes

- The rate of technology change in government departments

- The level of oversight KIM teams have on digital record holdings

5.3   Formats and technology in government

Over time, record formats and technology platforms change: therefore, an understanding of how a government department has produced and managed its born-digital information in the past helps the department to know how to look after its digital information assets now. How digital information has been handled over time also has an impact on how records are transferred to The National Archives and on preservation once they are in the custody of The National Archives. There is also the risk that, if departments do not have good knowledge of the record formats that they hold and actively work to care for them, formats could become obsolete and departments would be unable to access their digital information, resulting in records being lost.

---

[7] nationalarchives.gov.uk/information-management/manage-information/policy-process/digital-continuity/

Therefore, the review attempted to ascertain the types of digital record formats that exist and that have existed in government, the rate of technology change and the types of technologies in use. It also sought to understand the oversight and influence that KIM teams have over technology use in government departments as well as during technology change, and what impact technology change might have on records and information management.

Understanding formats and the types of technology in use in government departments will help The National Archives to offer advice and guidance to departments on how to manage their digital records over time, and on how to protect their records when going through a systems migration. It will also ensure that The National Archives can prepare its own systems to receive, preserve and present digital records.

In 2012-13, The National Archives conducted a survey of departments in which it asked departments to provide information on what types of digital information their systems held.

55% of departments surveyed reported having PDFs, and 50% reported having MS Project and Primavera formats. Most departments (40%) also reported that they had information in Microsoft Word format.

Both Primavera and MS Project are project management software applications that were used in departments in the past. The file formats from these applications are not readable in any other system and, therefore, if these types of files are selected for permanent preservation, The National Archives will need to advise on how to export the formats from these systems or decide how best to present the information.

It should be noted that the 2012-13 digital questionnaire only collected format information for records held in active business systems, not in legacy digital records. At Gateway 3 we will discuss the level of oversight government departments have over their legacy digital record holdings. For the purposes of this report, legacy records mean those that are not held in active business systems.

The longevity and usability of file formats is also predicated on how well digital records are migrated from older systems into newer ones, which is affected by the rate of technology change across departments.

5.4   The rate of technology change

Currently, government departments are undergoing rapid technology change due, in part, to policies encouraging departments to move away from large IT contracts and focus on working with small to medium sized enterprises. The Cloud First policy[8] that encourages departments to consider using Cloud service provision models has also fed into the technology change happening across government. At the time of our 2012-13 questionnaire, many government departments were considering a wide range of technological upgrades from mobile devices (60%), Cloud offerings (58%), to operating system (OS) upgrades (53%), as illustrated in Figure 1.2.

---

[8] Cabinet Office and The Right Honourable Lord Maude of Horsham, *Government adopts Cloud First' policy for public sector IT, 5* May 2013, accessed: 14 July 2015: gov.uk/government/news/government-adopts-cloud-first-policy-for-public-sector-it. David Thorpe, *How GDS is working to help small businesses*, 14 May 2013, accessed: 14 July 2015 https://gds.blog.gov.uk/2013/05/14/help-small-businesses/

Figure 1.2: Technologies being assessed for enterprise-wide use between 2012 and 2013 (Source: Digital questionnaire 2012-13)
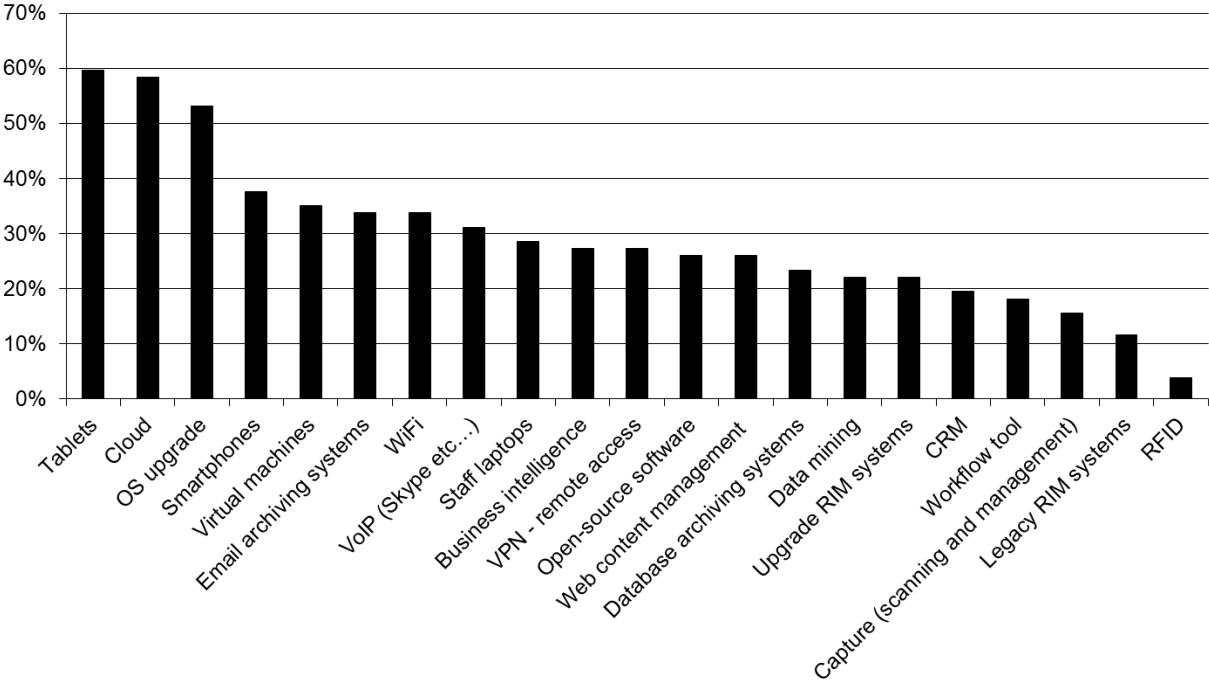
The Business Intelligence Review found that 81% of government departments surveyed had recently undergone a change in technology, were in the process of updating their technological platforms or were planning on updating their technology.

Figure 1.3: Technology change in government (Source: Business intelligence interviews 2014-15)
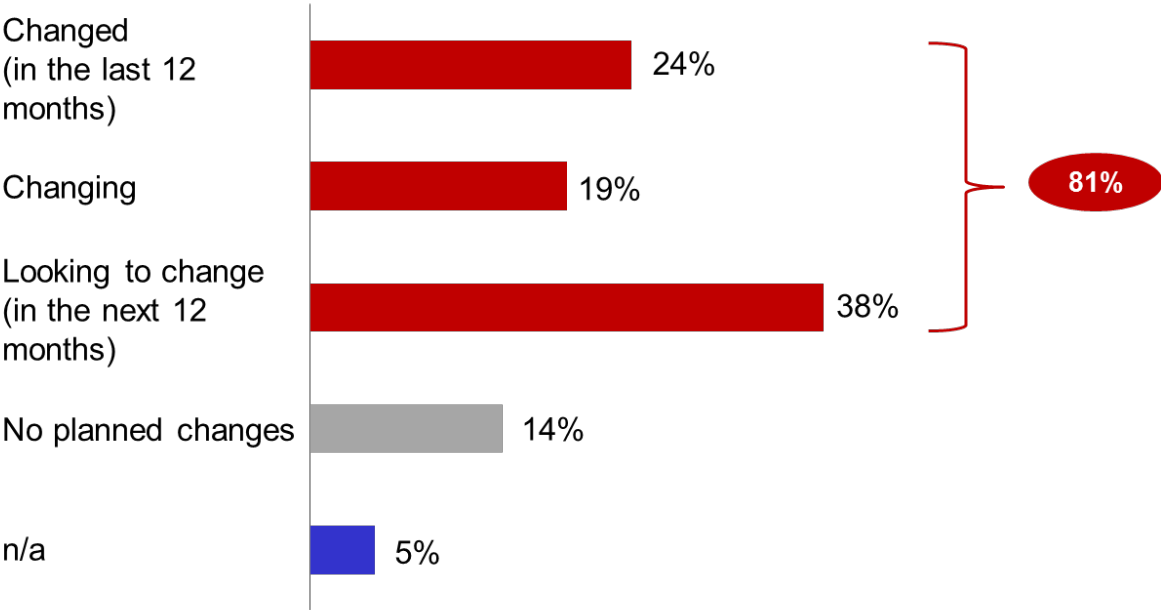
The rate of technological change is best exemplified by exploring the differences in information management platforms in use in government departments 2012-13 and those in use 2014-15.

It is clear that there have been significant changes in the types of information management systems used in the relatively short space of time between 2012-13 and 2014-15.

Aside from shared drives, the main platforms used in 2012-13 were Trim, Livelink, SharePoint and Meridio. However, in 2014-15 there was a dramatic reduction in the use of Meridio, Trim and Livelink, an increase in the use of SharePoint and an introduction in the use of Google Drive. At the time of this report, only Cabinet Office (CO) and DCMS are using Google Drive, but there are other government departments – e.g. Her Majesty's Revenue and Customs (HMRC) – that are starting to roll out this system as well[9]. The Department for Business, Innovation and Skills (BIS) has also rolled out Alfresco.

Cloud-based information management systems can allow employees in government departments to work with greater flexibility and collaborate more efficiently. However, there are information management challenges surrounding these systems that are becoming apparent. Without proper planning, migration to these systems can result in a loss of information – for example, a loss of crucial metadata such as creation date, title, creator and last modified date. It is also possible that system shortcomings could be overlooked when systems are being procured: for example, can these systems capture emails? This is significant because email is an increasingly important government record – more and more government decision-making is being documented over email.

5.5 Knowledge and Information Management (KIM) influence over technology change in departments

Knowledge and Information Management teams need to be able to monitor, influence and contribute to the technology change initiatives in their departments, as this greatly impacts their ability to meet digital continuity requirements.

Essentially, KIM teams should have sight of all the technology systems in place within a department to ensure that business records with long-term value are properly managed. They should

- know who is responsible for tracking all departmental systems

- be able to see what systems exist and can be accessed in departments

- know who is responsible for maintaining those systems and be able to ensure that the information is properly cared for

- feed into the procurement of new systems.

KIM teams should ensure that technology initiatives affecting the creation and care of digital records are designed according to The National Archives' advice and guidance,[10]

---

[9] Doug Boulton, 'HMRC to use Google Apps instead of Microsoft software- despite their past frosty relationship', *The Independent*, 6 June 2015, accessed: 14 July 2015: www.independent.co.uk/life-style/gadgets-and-tech/hmrc-to-use-google-apps-instead-of-microsoft-software--despite-their-past-frosty-relationship-10301932.html
[10] nationalarchives.gov.uk/information-management/manage-information/policy-process/digital-continuity/

which will ensure that departments are able to use and understand the records until they are transferred.

The review found that, while some KIM teams have the ability to feed directly into the procurements of IT systems – thereby making it more likely that applications can create and manage business records – this type of arrangement is rare. The review found that the ability to influence IT procurement partly stems from a strong working relationship with Technology departments. When such a relationship is not in place, information management systems may not meet KIM requirements or user needs.

If users are not considered when designing systems then there is a risk that they will not use the system. For example, if there is no easy way of getting emails into a system, such as a 'drag-and-drop' functionality, then users may find alternative areas to store emails or not file them at all. This results in inconsistent information capture and fragmented filing, making it difficult to fully document decisions, policy development and other business activities.

Information Asset Registers (IARs) allow KIM teams to keep sight of the types of systems that exist in their departments. An information asset is a body of information that is defined and managed as a single unit and has recognisable and manageable value, risk, content and lifecycle[11]. Examples of systems included on an IAR are information management systems and HR and financial databases.

The Business Intelligence Review found that the use and usefulness of IARs varied across departments and that the ability of KIM teams to see and contribute to the IAR also varied.

Figure 1.4: KIM Oversight on IARs (Source: Business Intelligence Review 2014-2015)
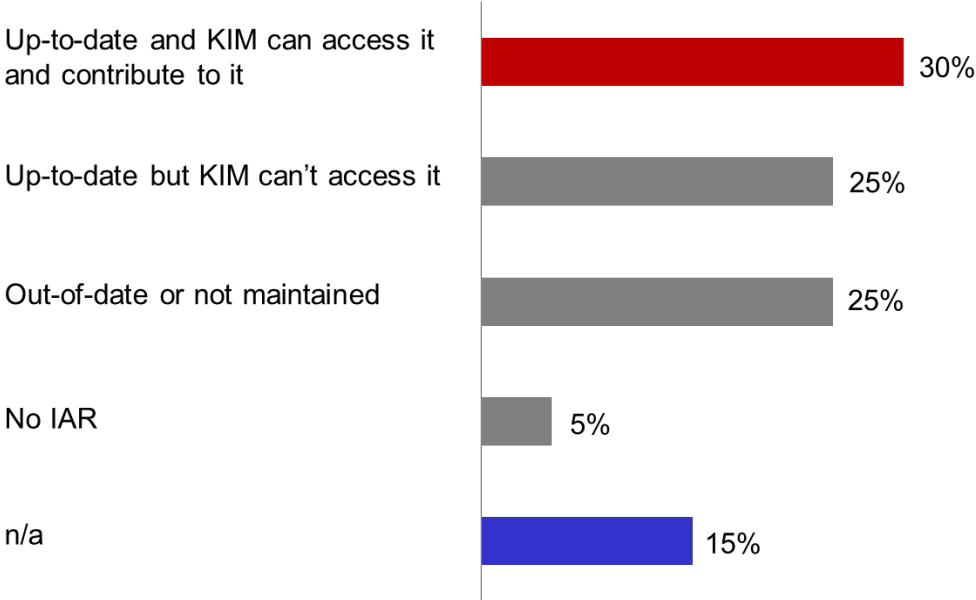
Out of 19 of the key departments surveyed, most KIM teams (55%) either did not have access to the IAR or could not contribute to it. This was either because there was no IAR or because another part of the organisation managed it or because it was not kept up to date.

---

[11] Definition from nationalarchives.gov.uk/documents/information-management/information-assets-factsheet.pdf

The absence of IARs, and the lack of KIM oversight where they do exist, may mean that departments do not know what systems exist in their departments. This could compromise their ability to manage their IT infrastructure and also increases the risk that multiple systems could be fulfilling the same function. Where KIM teams are unable to access the IAR, there is a risk that they will not be aware of important digital records and information residing outside of the information management system. This may result in digital information not being managed and even in digital information being lost.

KIM teams need to build strong relationships with their information technology departments in order to influence technology change within their departments. KIM teams also need to have an oversight of the whole of their digital holdings by using tools such as IARs. Having this overview of their information assets will allow them to understand what systems exist outside of any information management systems.

5.6   Recommendations and areas for further research

The National Archives needs to stay abreast of new technology developments in order to be able to guide and advise government departments in the care and maintenance of their records until transfer, as well as to be able to prepare and adapt preservation and migration approaches.

The business intelligence review, along with other data collected within the last three years, has given The National Archives a robust overview of the strengths, weaknesses, constraints and opportunities in the management and care of digital records.

In this changing landscape, it is important for government departments to ensure that digital records are actively managed through any systems migration and remain accessible and useable until transfer to The National Archives. For this to be achievable KIM teams need to

- know if they can open, work with and trust their information
- develop strategies to understand how born-digital records and their metadata can be moved from old systems into new systems
- build strong working relationships with information technology teams, information assurance teams and teams responsible for IARs.

The National Archives will also support government departments to help them fulfil their digital continuity responsibilities by

- staying at the forefront of the implementation of new information management systems
- carrying out regular IMA assessments
- conducting a review of the government digital landscape every two years
- influencing government procurement with Crown Commercial Service (CCS) around information management systems and, if possible, provide functional requirements that can be included in the Digital by Default Service Standard
- working closely with Government Digital Service (GDS) in areas that may affect the government technology offer, in particular areas such as Government as a Platform and the CTS

- refreshing existing guidance[12] on the need to protect born-digital records, targeting Chief Technology Officers (CTO) and supporting government departments in the development of digital continuity plans
- working with GDS on the development of a common user-needs framework for documents and records management.

6. Appraisal and selection (Gateway 1)

Appraisal is the process of distinguishing records of continuing value from those with no further value[13]. Records can possess different types or degrees of value to a public records body and this affects how records are managed and how long they need to be kept. This applies to records in all formats and media. The process of appraisal should help government departments and agencies understand which records are likely to have wider historical value and should therefore be kept indefinitely.

Selection[14] is a decision-making process that encompasses initial appraisal judgements and determines which records will be transferred to The National Archives or other approved place of deposit. The process of selecting records is primarily focused on determining which records hold a historical value to an organisation or wider society.

6.1   Department responsibilities relating to appraisal and selection

It is the department's responsibility to identify records of historical value and make selection decisions under the supervision of The National Archives. The appraisal and selection process should be done in a way that most efficiently and effectively allows for records of historical value to be identified. There are different methodologies for doing this work depending on the nature and resources of the department and the arrangement of the record holdings.

6.2   Areas the business intelligence review examined related to digital continuity

Under this Gateway the review examined

- how information such as digital records and emails is kept and disposed of, including the volume of information stored across different technology platforms (e.g. shared drives and information management systems)
- how social media is maintained and documented within government departments.

Appraisal and Selection is different in the digital environment from the paper environment because of the much more fragmented nature of digital record keeping practices. Digital records are often stored across multiple systems – for example, information management systems, shared drives and social media platforms.

---

[12] nationalarchives.gov.uk/information-management/manage-information/policy-process/digital-continuity/
[13] nationalarchives.gov.uk/information-management/manage-information/selection-and-transfer/appraising-records/
[14] nationalarchives.gov.uk/information-management/manage-information/selection-and-transfer/selecting-records/

It can be difficult to consider all the information as a whole across all the different systems and platforms to make meaningful decisions about what to keep. Also, digital records are not always kept in meaningful file structures and do not always have meaningful file titles: this adds to the challenge of determining which records have ongoing value. Accordingly, different approaches need to be used when appraising and selecting records in a digital environment. The approaches used now may even have to change, and in the future digital appraisal could differ even further from paper appraisal.

The business intelligence review sought to identify whether government departments had mechanisms for identifying records of enduring value and how evenly these were applied. It also examined email management and social media capture, as increasingly these mediums are, or are becoming, important documentary evidence for government activities. Finally, to fully appreciate the challenges of digital appraisal and selection and in light of the two previous points, the review attempted to scope the volumes of information across different information management platforms and look at how this affects the ability of government departments to carry out appraisal and selection in a scalable way.

6.3   Retention and disposal: what to keep and for how long?

Determining how long to keep records is governed by many different types of policies and procedures such as:

- What to Keep (WTK) schedules
- retention and disposal guidance/schedules
- Operational Selection Policies (OSPs)
- appraisal reports.

What To Keep schedules and retention disposal schedules and guidance help government departments determine what records to keep for business purposes[15], whereas OSPs and appraisal reports are tools used by departments and The National Archives to determine records of historical value that need to be selected and transferred to The National Archives. The focus of this discussion will be on WTK and retention/disposal schedules and their application within government departments as part of their wider KIM programmes.

The review found that the use of retention/disposal guidance and WTK schedules for digital records varied widely across government departments. According to the data available, a quarter of departments surveyed only applied their retention and disposal schedules to paper records and 15% were either revising their guidance for digital records or had no up-to-date guidance for digital records.

Half the departments surveyed tried to apply their retention rules across paper and digital platforms with varying degrees of success. Figure 1.5 provides a more granular breakdown of how departments have tried to apply this approach.

---

[15] Ministry of Justice. 'What to Keep'. *Retention and disposition schedules.* 15 December 2014. Updated 20 July 2015, accessed: 21 July 2015: gov.uk/government/publications/record-retention-and-disposition-schedules

Figure 1.5: Application of retention/disposal across paper and digital in 19 of the key government departments (Source: 2014-15 business intelligence review)
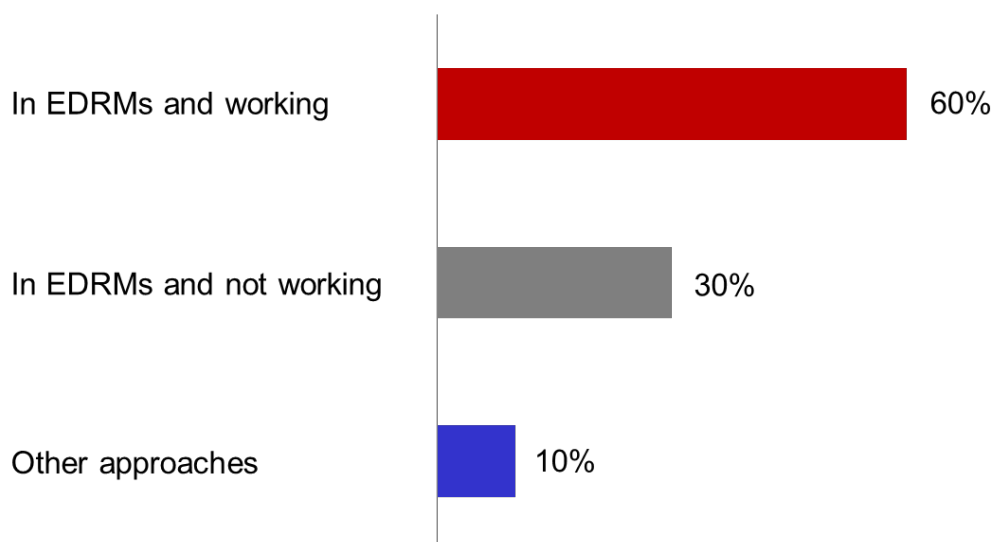
In Electronic Document and Records Management Systems (EDRMS) and working – this means that the length of the retention and disposal is programmed into their information management system and at periodic intervals reminders are sent to designated individuals to review the records.

The business intelligence review found that disposal decisions were not always tasked to KIM teams and often KIM teams identified business owners to review and authorise records for destruction. The reasoning behind this approach is that it is likely that the business owner will have a far greater understanding regarding what is of value to that team or department.

There are various advantages and disadvantages to this approach. Allowing business owners to review disposal decisions invests staff with a responsibility to properly manage and identify business records of ongoing value. However, this approach only works if there are strong governance structures in place that guide employees in the decision-making process. These should be complemented by regular staff training and an effective engagement programme.

It is promising that 60% of the departments surveyed had systems where retention and disposal functionality was working even if there remain some questions as to how consistently these processes are monitored by KIM teams.

In EDRMS and not working – the departments in this category had an in-house information management system but the functionality that allowed them to automate the retention/disposal function was either not working or not programmed into the system.

This can create many challenges when it comes to the appraisal and selection process as there may be vast amounts of information that must be assessed at the time of transfer, which could put pressure on limited personnel and financial resources. Furthermore, large amounts of information can affect the efficiency of search functionalities that are built into information management systems. Searches may return vast amounts of records that users must filter to identify information that they need.

In one instance, a department had retention/disposal schedules but no information management system. They managed their records in a certified shared drive environment. The KIM team could control how users filed their records and how information was organised. They used the guidelines set out in their retention/disposal schedules to govern both their paper and digital records retention, although identifying digital records for disposal was difficult for this department.

Most retention/disposal and WTK schedules only govern records created in information management systems or information management environments and do not always extend to information held in collaborative areas such as shared drives or on social media platforms.

6.4   Email management

Emails are an integral part of government administration as they make up what was formerly correspondence or inter-departmental memoranda documenting exchanges or decisions pertaining to a particular topic. Under the Civil Service Code[16], Civil servants have an obligation to keep accurate official records. This means that departments need to

- **define** clearly which emails need to be kept for business or historical value

- **communicate** simply and often to users the rules about what emails to keep

- **keep** emails with related digital information in a shared corporate information management systems

- **limit** what users can keep in personal email accounts by the use of:
  – email account quotas OR
  – automatic deletion after a set period of time

- **not use** any system that changes email formats into ones that are no longer usable (e.g. a text file)[17].

The approach to the capture and management of emails varies significantly across departments. Some departments put limits on the size of inboxes where others apply no limits.

The importance of an information management system that is easy to use – for example, one with drag-and-drop functionality – appears to impact significantly on the ability of departments to capture emails. An easy-to-use system can contribute towards a systematic capture of email, but there is the risk that users can indiscriminately bulk export emails from inboxes into information management systems without careful consideration for what should be captured. There is also the possibility that, regardless of how easy the system is to use, users still may not take the time to file their emails. The review found that in some departments there was a need for cultural change as well as a system that is easy to use. For example, one department had an EDRMS with drag-and-drop functionality but they also had an email server that contained ½ billion emails.

---

[16] gov.uk/government/publications/civil-service-code/the-civil-service-code
[17] nationalarchives.gov.uk/information-management/manage-information/policy-process/managing-email/

Further findings from the business intelligence review indicated that 50% to 70% of the contents of some departments' information management systems are comprised of emails. Where departments do not have an information management system with a drag-and-drop functionality, or where records are managed in a shared drive environment, there is a serious risk that information could be lost. The review also found that, if users do not have the ability to save emails to folders easily, email servers tend to suffer from 'email bloat', meaning that they are very large. Numbers given by government departments ranged up to 190 TB held in email servers.

The business intelligence review also found that departments have different retention rules regarding their emails. Some have automatic deletion policies with periods ranging from three months to two years. The risk where emails are automatically deleted from email servers is that, if users have not been periodically moving important emails to another environment to ensure they are captured, they may be lost.

The 2012-13 digital questionnaire indicated that 42% of government departments surveyed had automatic email deletion policies. Only 30% of those departments archived their emails and 22% of those departments had an integrated drag-and-drop functionality within their information management system.

Not only does the capturing and management of emails pose challenges for departments, but searching emails holdings can also be problematic. The 2012-13 digital questionnaire found that 64% of departments surveyed did not have the ability to search .pst files (Outlook email messages) in either local or networked environments. Departments need such capability to be able to engage with large Inquiries where they will need to demonstrate the completeness and accuracy of their records across different digital platforms.

6.5   Collaborative environments and social media

In recent years there has been a greater uptake in the use of collaborative platforms and social media in government. In particular, policies set out by GDS are encouraging government departments to utilise collaborative environments to improve efficiency and information sharing. As well as social media such as Facebook and YouTube, collaborative platforms such as Trello, Smartsheet and Yammer are also in use. This is a promising step forward to enable better information sharing across government; however, departments need to continue to ensure that digital records of business value are captured and kept. Some departments such as the Ministry of Justice (MoJ) and Department of Health (DH) have developed policies governing the use of collaborative environments for work and the need to save records of corporate value.[18] Capturing records from these work environments may or may not be feasible, as some may not have export capabilities or may provide documents in proprietary file formats that may not be able to be interpreted by any other software.

The National Archives and GDS will work together to best advise government departments on the use of collaborative platforms to ensure the capture of business records.

---

[18] Stephen Hale, 'Use of Third Party Tools' *Digital Health- Department of Health*, 9 April 2013, accessed: 22 July 2015 http://webarchive.nationalarchives.gov.uk/20140522191340/http://digitalhealth.dh.gov.uk/use-of-third-party-digital-tools/ Ministry of Justice, 'Use of Internet Tools' no longer published, accessed: August 2014

Figure 1.6: Use of social media platforms in government (Source: 2012-13 digital questionnaire)
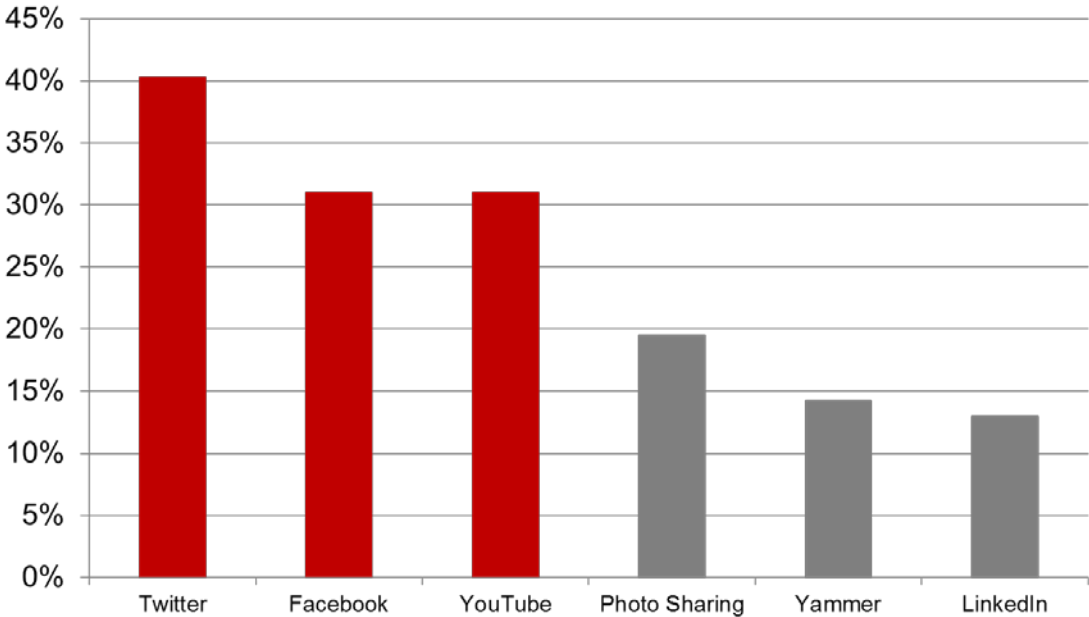
Unsurprisingly, the most popular platforms were Twitter (40%), Facebook (31%) and YouTube (31%).

There are broader issues than just capture to consider when dealing with social media, such as whether or not it is considered of value as a business record, and its value as a historical record of government processes.

The 2014-15 business intelligence review only received data from half the respondents on their social media management strategies. Of the responses received, only 10% of government departments reported that they were actively capturing social media content or that they required staff to ensure that output government created on social media was captured. 30% had a medium level of capture and 60% were not capturing social media at all, although some were looking at developing policies around social media capture.

Figure 1.7: Capture of social media (i.e. Twitter, Facebook) by 10 government departments (Source: 2014-15 business intelligence review)
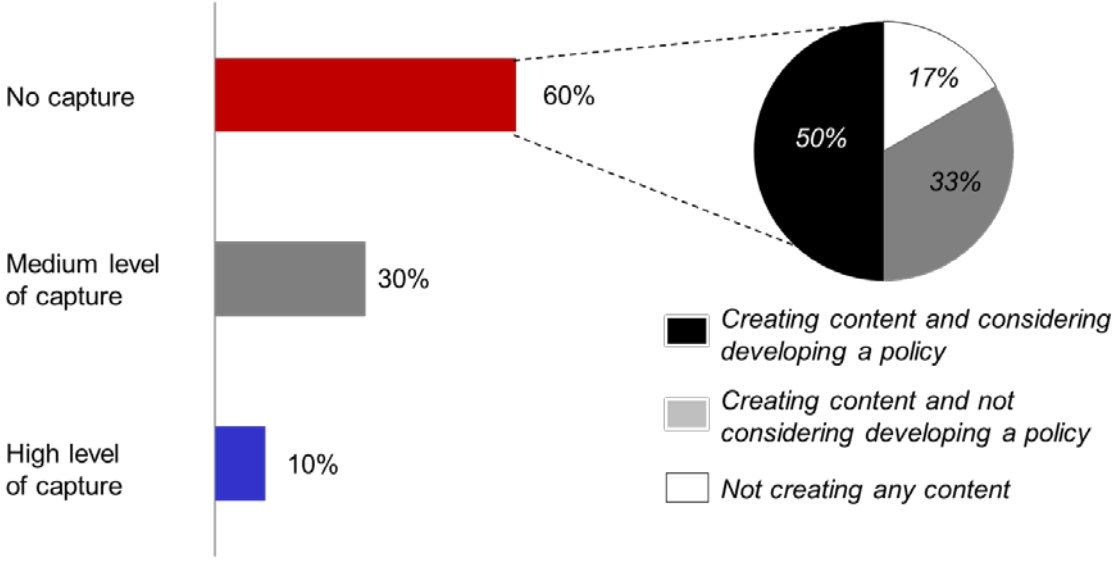
As the use of collaboration platforms and social media becomes more popular in creating, sharing and disseminating information, the types of platforms being used by government departments will change.

6.6   Information management platforms: EDRMS vs shared drives

The business intelligence review found that two thirds of data held by 11 out of 19 of the key government departments interviewed is on shared drives.

In some government departments, for every TB of information held in an EDRMS, there are approximately 10TB of data in shared drives. Even when an EDRMS exists, shared drives often exist in parallel. Shared drives often have no type of record-keeping controls in place to limit or monitor filing. The following chart illustrates the proportion of departments that have shared drives and EDRMS.

Figure 1.8: Percentage of data from 11 out of 19 key government departments held in EDRMS vs shared drives (Source: business intelligence interviews 2014-15)
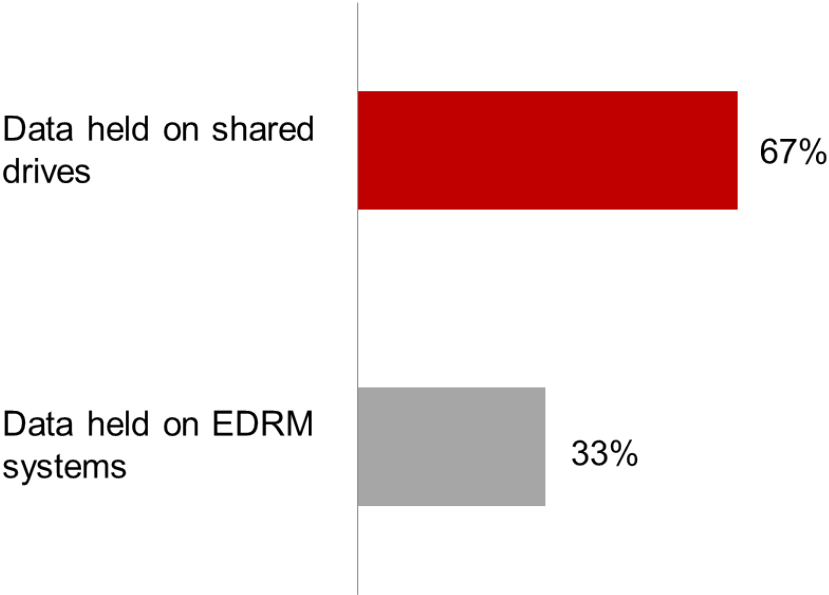
Storing data in parallel areas risks duplication across platforms: this can have cost and resource implications when it comes to appraisal, selection and sensitivity review. There is also a risk that records stored in shared drives will never be moved across to corporate information management systems.

Shared drive environments do not permit the same level of central record-keeping control as information management systems such as EDRMS. The business intelligence review found several instances in government departments where shared drives had no file structure applied to them and were comprised of a running sequence of records. Many of these records also had poor file titles, making it difficult to locate and understand the contents of these records. Poor naming conventions, lack of file structures and limited search capabilities make the retrieval, appraisal and selection of records of long-term value at a macro-level at scale difficult for government departments.

6.7   Recommendations and areas for future research

The National Archives needs to stay at the forefront of technological advances in order to be in the best position to advise government departments on the capture of their digital records. The National Archives could also consider what future solutions it may want to develop to help government departments in this area, as well as the guidance it offers to support government departments in the management of their digital records.

Appraising and selecting born-digital records at scale, 20 years after their creation, will be a challenge for government departments. In light of these challenges, The National Archives has already begun exploring different types of technology to assist with the appraisal and selection process for unstructured information (i.e. shared drives). Technologies such as Latent Dirichlet Allocation, Latent Semantic Indexing and

relational databases as well as eDiscovery tools are being tested to explore how they can help to cluster and categorise records in order to extract meaning – enabling selection and appraisal decisions at a high level rather than through document-by-document review.

These technologies are particularly interesting for

- **understanding the information at a high level** (e.g. volume by dates of creation, split by format)

- **reducing the amount and prioritising the information to review** by excluding exact duplicates, near duplicates and non-meaningful files or formats (e.g. calendar invitations, out-of-office files)

- **clustering the information** automatically or against pre-set categories.

The National Archives will publish a report exploring the work it is doing in this area in February 2016. The report explores how existing technologies can assist government departments in both born-digital appraisal and selection and sensitivity review and how this technology-assisted review could prove to be just as accurate and consistent, if not more so, when compared to a human review.

Moving forward, The National Archives will work collaboratively with government departments, GDS and CCS to think about developing a cost-effective, flexible shared-service approach. In particular, it needs to consider how government departments will be able to carry out cross-platform appraisal and selection, bringing together information from email servers, information management systems, data management systems, social media and other platforms. This is to enable an assessment of the totality of the records that have value, not simply a segment of records that are contained in an information management system or shared drive.

In answer to the challenge of capturing emails, The National Archives should examine other approaches and consider recommending alternative methods such as the US Capstone policy. The Capstone approach consists of retaining the email collections of senior officials as a method for documenting decision-making and organisational operations.[19] There are also data-analytics tools that can help identify potentially important emails or email collections to ensure the capture of those that have corporate value. The Capstone approach, combined with using tools to appraise, select and make sense of email collections, could ensure full capture of email collections as well as ensuring that, once captured, the collections could be given context and be made available to the public, subject to sensitivity review.

The questions that The National Archives should explore around email management and social media as well as collaborative platforms include:

- How can we identify emails with long-term value, at scale? Would the US Capstone policy on email capture be something to consider for UK government departments?

- Are there ways to bring meaning to email collections in order to make more informed appraisal and selection decisions?

---

[19]https://www.archives.gov/records-mgmt/bulletins/2013/2013-02.html

- How should we capture and present email attachments? How can we ensure that members of the public can find relevant material in the attachments in the way emails are catalogued?

- Should social media be considered a public record?

- What is The National Archives' recommended approach for capturing social media?

The National Archives is already exploring collaborative research with academics, looking at approaches to managing emails in the long term. Where there is no natural lead, The National Archives could look at spearheading its own research.

7. Sensitivity review (Gateway 2)

Sensitivity review is the process by which government departments and agencies, under the supervision of The National Archives, identify material that

- should be *retained* by the creating department, as the records are too sensitive for immediate transfer

- should be *transferred* as *closed*, as Freedom of Information (FOI) exemptions apply

- can be *transferred* as *open*, as no FOI exemptions apply.

The Advisory Council on National Records and Archives validates the applications for the retention and closure of records, along with the closure exemptions applied, including the closure periods.

Digital sensitivity review presents particular challenges – for example, being able to accurately identify sensitivities in records at scale and ensuring consistencies of sensitivity review across formats (i.e. paper and digital).

The business intelligence review found that only a small number of government departments were currently thinking about and tackling the challenges of digital sensitivity review. This is due, in part, to departments needing to accelerate the review of paper records in order to meet the change in the timing for transfer from 30 years to 20 years.

The following section examines the sensitivities present in UK government departments, based on an assessment of applications to the Advisory Council, and also examines the state of readiness of government departments to tackle the challenges of digital sensitivity review.

7.1 Department responsibilities relating to sensitivity review

Under the Section 46 Code of Practice[20], departments are required to determine the access status of their records before transferring them to The National Archives.

---

[20] https://ico.org.uk/media/for-organisations/research-and-reports/1432475/foi-section-46-code-of-practice-1.pdf

7.2 Areas the business intelligence review examined relating to sensitivity review

Under this Gateway, the review endeavoured to

● understand how prepared government departments are to undertake digital sensitivity review

● identify areas where government departments may require greater assistance or support from The National Archives during the digital transfer process.

7.3 Sensitivities at stake in UK government departments

Almost all the government departments surveyed as part of the business intelligence review cited reviewing born-digital records for sensitivity as one of the major challenges ahead of them, prior to transferring their records to The National Archives.

In order to be able to better understand the sensitivity review challenges facing government departments and to develop targeted solutions tailored to government departments' needs, the business intelligence review looked at the kinds of sensitivities in records across government departments. It did this by analysing all the Advisory Council applications for paper record transfers closed under the FOI Act at the point of transfer or following an FOI request between 10-02-05 and 30-04-14. This examination allowed the review to identify the most common FOI exemptions used by government departments when closing records.

This analysis of the applications for closure by key UK government departments found that 75% of closed items and pieces related to personal sensitivities – sections 38, 40 (2) and 41 of the FOI Act. Figure 1.9 details the proportion of applications by exemptions made by the key 21 government departments to the Advisory Council.

Figure 1.9: Split of closed items and pieces by type of exemptions - key 21 government departments (Source: The National Archives data)
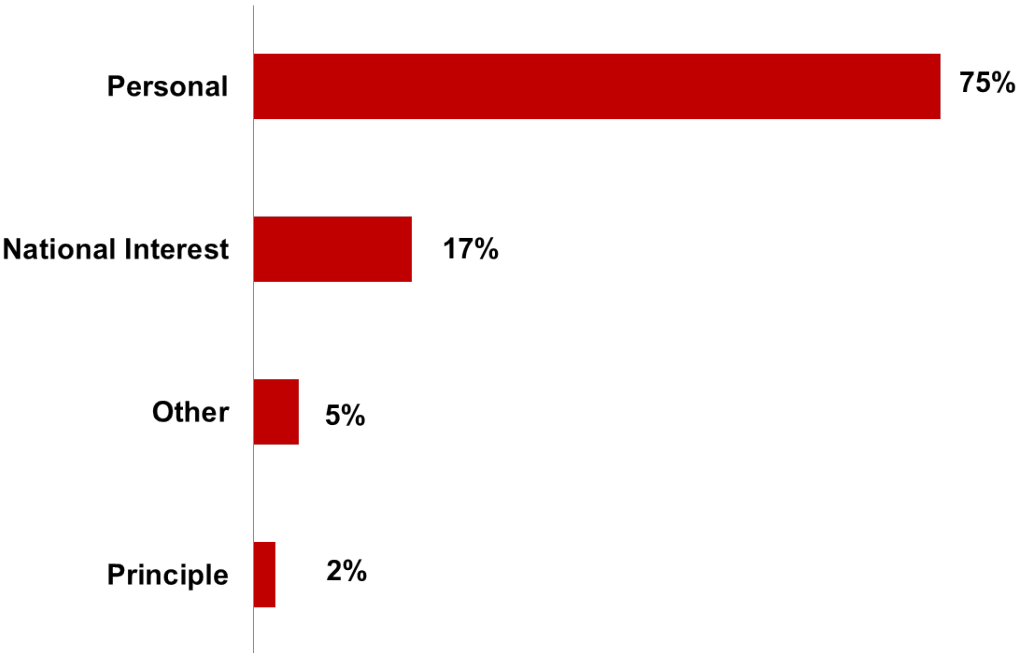
Personal – relates to personal sensitivities detailed under the Data Protection Act (DPA)[21] and includes information such as names, National Insurance numbers, telephone numbers, religious affiliations, dates of birth (section 40 (2)) as well as health and safety (section 38), and information provided in confidence (section 41).

National interest – includes exemptions such as international relations (section 27), national security (section 24) or the economy (section 29).

Principle – this relates to matters dealing with the development of policy (section 35), communications with Royal Household (section 37) or legal privilege (section 42).

Other – this covers a wide range set of other exemptions such as court records (section. 32), law enforcement (section 31) or environmental information (section 39).

For a more detailed explanation of the exemptions and their applications, see the published guidance.[22]

In order to prioritise solutions and break down the challenges of sensitivity review, this analysis sought to determine whether any departments only made applications for certain exemptions. Using the applications for closure, we determined that government departments can be clustered into two groups:

- Government departments **with almost exclusively personal sensitivities**.
- Government departments **with 'blended sensitivities'**. By 'blended sensitivities' we mean departments that have made a broad range of closure applications under different exemptions. These departments often have a high proportion of closed items relating to 'National interest' (section 23, 24, 26, 27 (1), 27 (2), 29 of the FOI Act).

The analysis of the applications for closure under personal data exemptions highlighted that, for 12 of the key 21 departments, more than 80% of the exemptions they apply for pertain to personal data. In some instances, there are government departments where 100% of their closures are solely related to personal data. We have broken down our definition of sensitivities according to these categories in order to begin developing strategies and methodologies for addressing digital sensitivity review.

Given that 75% of government departments are making closure applications under personal sensitivity exemptions, The National Archives decided to test several commercial data analytics software applications to see if they could help in the identification of personal sensitivities in born-digital records at scale. Our software trials only test how the software can identify personal sensitivities, as personal sensitivities tend to be more structured in nature and are potentially easier to characterise. 'Blended sensitivities' are much more difficult to identify as they tend to be more contextual, and current data analytics software does not have the capability to identify these sensitivities with a high level of accuracy. The National Archives aims to have completed the software trials and publish the findings at the beginning of 2016. Based on these trials, The National Archives will develop methodologies to assist government departments in addressing this component of the digital sensitivity review challenge.

---

[21] *Data Protection Act 1998* (UK), s 2, accessed: 27 July 2015, legislation.gov.uk/ukpga/1998/29/section/2
[22] Ministry of Justice, *Exemption guidance for freedom of information (FOI),* 3 October 2014, accessed: 27 July 2015, www.justice.gov.uk/information-access-rights/foi-guidance-for-practitioners/exemptions-guidance/

7.4 Digital sensitivity review in government departments

At this point, only 33% of the key government departments are actively trying to address the digital sensitivity review challenge. Figure 1.10 provides an overview of the responses from the business intelligence review around the topic of digital sensitivity review.

Figure 1.10: Percentage of 19 of the key government departments that have started to develop strategies to perform a digital sensitivity review of their records (Source: business intelligence interviews 2014-15)
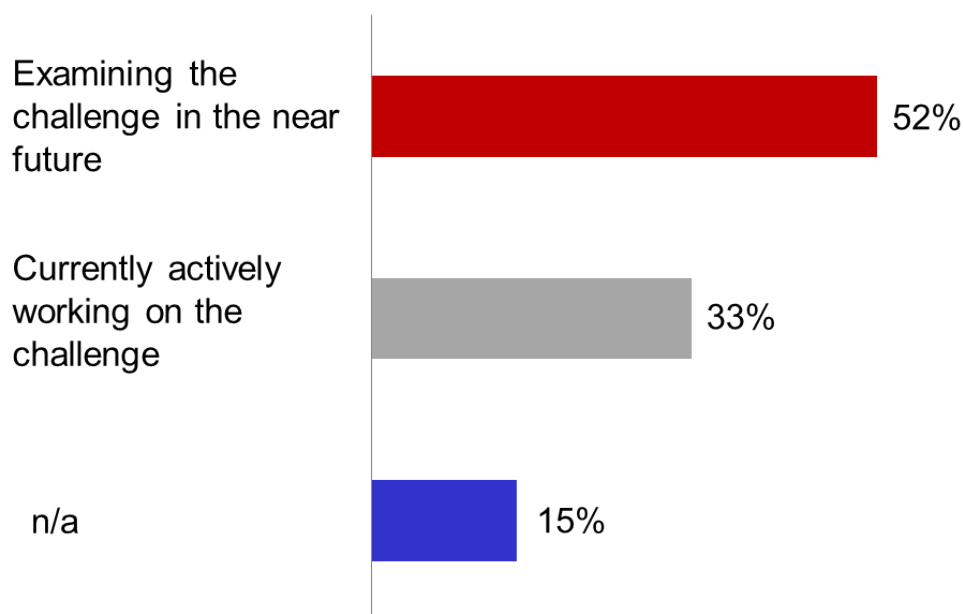
Based on The National Archives' early digital transfer pilots performed as part of the DTP, we know that the challenge of digital sensitivity review is compounded by the sheer volume of digital information as well as its discoverability.

The current economic climate, and reduced resources, makes these challenges harder to address for many departments. It seems that departments are aware of the digital sensitivity review challenge, but many have not yet fully explored how to address it.

7.5 Recommendations and areas for future research

The National Archives is already working to find solutions. In particular, our software trials have started exploring the suitability of commercial tools to help government departments tackle the challenges of digital sensitivity review at scale.

There are existing tools, technologies and research that could potentially support government departments in reducing and prioritising the amount of information to review. These tools include e-Discovery tools, Latent Semantics Indexing, relational databases, Latent Dirichlet Allocation. The findings from our initial software trials show that, although there is no silver bullet, some of these tools may be promising in helping to

- identify areas of known, text-based sensitivities such as email addresses, NI numbers, bank account details, phone numbers, and known names in a systematic way

- automatically redact this sensitive information

- sort, reduce and prioritise information to be reviewed.

It is important to note that the challenge of sensitivity review is not just one that can be solved by a powerful search engine. Sensitivities are also based on context. The National Archives is currently working with the academic community and with other government departments to find scalable solutions for the long term.

The National Archives has identified the following areas that will need to be investigated further to help government departments with the challenge of sensitivity review:

- How to find all types of sensitive information at scale contained within records and their metadata

- How to effectively plan resources for sensitivity review at scale in the future

- How to address the challenge of reviewing material on screen for a long number of hours

- Existing technical solutions that could help government departments with their sensitivity review, in a defined level of risk

There are also several other issues that government departments, in tandem with The National Archives, will also need to examine, such as

- the skills and capacity that digital sensitivity reviewers will need

- how best to manage retained records and link them to transferred records.

8. Transfer (Gateway 3)

This is the process by which government departments physically transfer their born-digital records and their metadata to The National Archives. This process has several technical requirements to ensure that the records and their metadata are transferred to The National Archives securely.

The business intelligence review sought specifically to understand the challenges and obstacles that could prevent government departments from meeting their transfer requirements under the PRA, and what challenges departments might face in meeting The National Archives' technical requirements.

8.1 Departments' responsibilities relating to transfer

Departments are required to ensure that their records are transferred to The National Archives.

8.2 Areas the business intelligence review examined relating to transfer

What are the major challenges that government departments face in preparing for digital records transfer? They include

- understanding their digital holdings and what date to use as a trigger date for transfer
- the challenges of outsourced IT
- the skills and capacity needed for the digital transfer process.

8.3 Knowing what digital records are held and compliance with the PRA

Departments need to know what they hold in order to understand what needs to be transferred; in particular, they need to understand the age of their digital records. This will not only assist them with their digital continuity planning but it will also help them plan their digital records transfers. At present, few departments have an overview of their holdings and can only estimate their projected transfer dates, at best.

This particular issue was highlighted when The National Archives sent out a digital questionnaire in April 2015. The questionnaire asked two questions:

- What is the date of your oldest born-digital information?
- What is the estimated volume of the earliest born-digital information you hold, based on the year identified in the first question, in that year only?

During this exercise, the DTP team realised that most government departments had little or no sight of their complete digital legacy holdings – something that was also reinforced by the business intelligence review.

This poses a problem for departments as they do not know when their digital records are due to be transferred to The National Archives. This underscores a wider issue of what constitutes compliance under the PRA when it comes to digital records transfers.

Many departments are using the end of their print-to-paper policy as a marker for their digital records transfers. Print-to-paper policies designated the printed and filed record as the authoritative record over the digital one. However, not all departments had these policies and many were applied unevenly. Currently there are departments that still have print-to-paper policies, but these policies are not always being followed. Therefore, print-to-paper policies are unreliable as a marker for digital transfers.

Given the uneven application of print-to-paper policies, it would be best for The National Archives to begin taking digital records from government departments based on the date of their earliest born digital record. This requires departments to have surveyed their digital records to find out what their earliest digital information is.

8.4   The challenge of outsourced IT

Outsourced IT means that government departments' IT infrastructure – such as the servers that host their software and sometimes their IT Helpdesks – are contracted out to a third-party IT provider. While this saves on infrastructure costs, it can actually cost government departments significant amounts of time and resources to try to introduce new software onto their systems. Part of the reason for this is that many of the IT contracts were negotiated over five years ago and did not factor in technological advances or the need to transfer digital records to The National Archives. The business intelligence review found that 76% of 19 of the key government departments surveyed had outsourced IT structures. Figure 1.11 provides the distribution of in-house versus outsourced IT.

Figure 1.11: Percentage of 19 of the key government departments who have outsourced IT (Source: business intelligence interviews 2014-15)
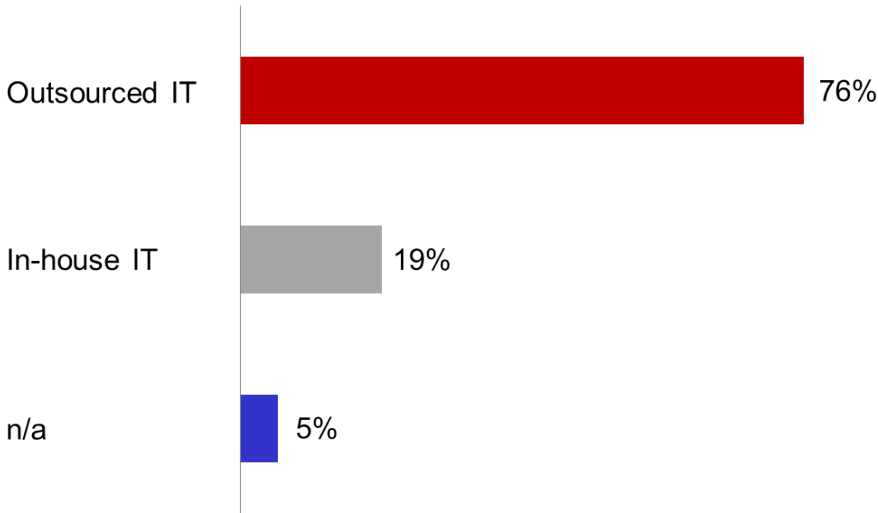
Outsourced IT poses an important challenge in transferring digital records because The National Archives requires government departments to download three pieces of software that are used at different points in the transfer process:

- **DROID (Digital Records Object IDentifier)**[23] – This is used to identify file formats and volumes. This is necessary so that the records can be ingested into our Digital Records Infrastructure (DRI). It is also the piece of software that creates some of the mandatory metadata about the digital records that becomes the basis for cataloguing digital records.
- **CSV validator** – This ensures that the information input into the digital records cataloguing template is technically correct.
- **Teracopy** – This is used in appraisal and selection (Gateway 1) to prevent any alterations to the digital records during the transfer process. This is particularly helpful in protecting against changes to the last modified date, which is the date by which The National Archives calculates closure (providing the last modified date is reflective of the contents of the record)

The business intelligence review and findings from the initial digital transfer pilots have indicated that impact assessments and beta testing prior to the full installation of the software can take from four weeks to 18 months.

There are also considerable financial implications for departments – each piece of software can cost up to £10,000 to test and install onto their networks. The Government Digital Service is trying to move departments away from being tied to long-term outsourced contracts with large IT providers, preferring instead that they retain a component of their services (i.e. helpdesk) in-house but continue to outsource infrastructure to small to medium-sized organisations, especially Cloud-based companies.

When planning the renewal of IT contracts, departments should ensure that they consider the requirements of transferring digital records to The National Archives.

The National Archives is proactively working to ensure that the software needed for digital records transfers onto the CTS approach developed by GDS. The aim is that government departments will avoid protracted and expensive software testing processes through a consistent GDS approved approach.

8.5   Digital Skills and Capacity

While government departments embrace new digital ways of working and move into exciting areas such as data analytics, there is a need to develop new skills and grow a qualified pool of individuals within government who have the required digital skills. When it comes to digital transfer, Gateway 3 could be one of the most challenging for departments as it differs from the paper process and requires a number of technical skills.

There is a need for greater technological proficiency among sensitivity reviewers. They need the technical skills to be able to work with new software (i.e. data analytics and eDiscovery) as well as being able to embrace new approaches to working with digital records.

Our pilot transfers and our first training sessions have indicated that there is a need for new technical skills within government departments in order to be able to cope with the

---

[23] nationalarchives.gov.uk/information-management/manage-information/policy-process/digital-continuity/file-profiling-tool-droid/

digital transfer process. These include improved Excel skills as well as the ability to understand and troubleshoot the error messages from the CSV validator – these skills do not necessarily currently exist within the KIM teams in government departments.

The question of skills and capacity is not just restricted to the digital transfer process. As previously discussed, digital records have wider implications on information management and digital continuity within departments. The breadth and complexity of digital records is not likely to diminish nor is the difficulty of trying to ensure they remain available and useable until transfer to The National Archives. Knowledge and Information Management teams within departments need advanced digital skills to give them the ability to adapt and think innovatively about digital information management solutions.

8.6   Anticipated volumes of transfer

Although KIM teams need to develop capacity, skills and resources to support the digital transfer process, The National Archives also needs to ensure that it can support the ingest of digital transfers.

The first business-as-usual digital transfers are due in 2016 from approximately 16 government departments. In the digital questionnaire, we asked government departments for approximate volumes for their first digital records transfers (based on their earliest born-digital information). These volumes were provided pre-appraisal and selection and therefore net volumes could be up to 40% or 50% less; but, based on this, in 2016 we are expecting up to 5TB of information.

8.7   Recommendations and areas for future research

Having an understanding of digital record holdings is vital to understanding not only when digital records need to be transferred but how they should be managed to ensure their availability over time. Outsourced IT has been identified as a major challenge while working on our digital transfer pilots. The National Archives will continue to collaborate with GDS and CCS to ensure that the software required for transfer is included in the CTS approach and the future digital procurement framework, which is being developed by CCS.

The question of skills and capacity is another important area for consideration as transferring born-digital records requires new, technical skills that can pose a number of challenges to existing records management teams in government departments. As we move forward, The National Archives needs to help government departments to identify minimum skills requirements for working in the new digital KIM profession.

The National Archives needs to do more research and development of DROID so that the digital transfer process is more easily scalable.

9. Conclusion

Digital records are changing what it means to be an archive and, although they present a number of challenges, they also present an exciting opportunity. The National Archives is one of the most accessible archives in the world, with open born-digital records available for free to view and download from anywhere in the world.

Departments need to work with The National Archives to fully understand their digital record holdings. Strong relationships need to be built in departments between records management, KIM and IT teams in order for the digital transfer process to be a success. With the support of The National Archives, departments need to be clear about their digital continuity requirements in order to be able to safeguard records in their custody prior to transfer.

Resources need to be allocated to tackle the challenges of digital sensitivity review and departments need to start to address their particular issues relating to digital sensitivity review. A collaborative approach of sharing knowledge and information across government will help departments to explore the best tools and technologies to assist in some of the digital challenges.

Across government the question of digital capacity and skills needs to be addressed. Many teams, such as KIM and records management as well as sensitivity reviewers, will need new digital skills in order to be able to successfully appraise, select, sensitivity review and transfer records to The National Archives.

The National Archives could also address the need for new research and guidance in the area of digital transfer in order to support government departments in the transfer process. This includes investigating current technologies that may be able to assist departments in the areas of appraisal and selection and digital sensitivity review.

The National Archives is committed to working on innovative solutions as well as working collaboratively with departments to embrace the challenges and opportunities of digital records. Sharing knowledge and ideas and exploring tools and technologies government-wide will be an important part of conquering the digital future.