

Meeting the threat to Portsmouth City Council's digital records

John Stedman, Records Manager & Archivist
Mike Greenslade, Information Security Officer

22 Dec 2017

1 Executive Summary

- 1.1 The city's electronic records are under threat from the upgrading of software and hardware and the consequent obsolescence of files created in older formats, and from changes to council responsibilities, staff structures and the normal wastage and replacement of staff. The council can be confident that, within 20 years, without action it will no longer be able to access thousands of records that it has legal or business reasons for keeping. It is not currently meeting the requirement in its Records Management Policy to "ensure that the records will be maintained in a format allowing them to be read regardless of technological change".
- 1.2 The council therefore needs a Digital Records Preservation Policy. The key action will be to establish a 'virtual' digital repository, similar to the Modern Records Section, to manage those digital records which it wishes to keep for over 10-20 years. Such a repository would:
 - maintain a watch on the viability of electronic records,
 - migrate records into new formats when threats to old ones emerge,
 - document this activity as demonstration of the records' reliability and legal admissibility,
 - preserve them from unauthorised amendment,
 - keep intellectual control by cataloguing what is being preserved,
 - provide reliable copies when needed.
- 1.3 A working group should be set up to determine the council's digital preservation requirements, decide on the most effective approach, estimate the resources needed and write an appropriate policy. Its membership should be drawn from the IT and Library & Archive Services, with the support of Information Governance.

2 Introduction

- 2.1 Electronic records are doomed to technical obsolescence and will become unreadable. They are vulnerable to accidental loss and deliberate amendment or destruction. Nevertheless, the long-term outlook for Portsmouth City Council's electronic records is unplanned and uncertain, despite the commitment in its Records Management Policy to "ensure that the records will be maintained in a format allowing them to be read regardless of technological change". This report draws attention to the implications of current inaction and suggests a path to resolving the potential problems. It outlines the principal threats to electronic records, discusses possible courses of action and concludes that the council needs to create and implement a Digital Records Preservation Policy. It expects that creating a digital repository will be the most effective approach. It proposes that a working party be created to define the council's requirements, suggest a staff structure to manage the repository and choose appropriate software to manage the records.

3 The current situation

- 3.1 The city has nearly 12 million documents on its W: drive and many more on other drives. Most record actions and decisions by the council or its customers, or information the council needs to do its work. They are in a variety of formats - MS Word, Excel spreadsheets, jpegs and tiffs, CAD plans, pdfs, to name a few; there are also documents in several electronic document management systems, such as EBS and Swift. Of these it needs to retain a small proportion permanently (eg, the minutes of council committees), others for up to 100 years and longer (eg, adoption and fostering records, tree preservation orders, building construction and management records), while there are more which should be kept for over 20 years (staff records, long-duration contracts under seal, planning documents). In addition, there are thousands of 'born-digital' documents that the council will choose to keep as an historical record and extensive digitisation projects carried out by museums and archive department staff, volunteers and in partnership with Find My Past and the Heritage Lottery Fund, representing substantial investment in time and money.
- 3.2 These are records that the council is obliged to keep by law, or wishes to keep as its corporate memory, allowing better decision-making and creating a sense of continuity and community. Losing them would badly affect the council's reputation as a competent and trustworthy representative of its citizens and would sap staff morale. Nevertheless, no provision is being made, beyond the basic one of keeping backups and restricting folder access

to particular teams, to ensure that these records remain readable for the periods they have to be retained, secure from unauthorised amendment or deletion, and demonstrably authentic if required for defence of the city's interests.

4 The threats

4.1 Digital records are fragile:

1. They are liable to corruption - the 'bitstream' of ones and noughts that make up a digital file can decay, and damage to even one bit may render a file unreadable.
2. Servers, on which most of PCC's records are stored, are relatively robust, but not infallible, while other media are unproven and prone to failure.
3. Software in use now, such as the Microsoft Office suite, may be replaced rendering it difficult to read records written with those programs. Within the last 25 years PCC has used and abandoned Polo as an e-mail system and Amipro for word processing; documents created in these packages cannot now be accessed without significant effort and cost.
4. Even without adopting different software programs, each time hardware or software is upgraded it becomes less likely that records created in older versions of software or operating environments will be accessible. 'Backwards compatibility' in software is not guaranteed; in some CAD programs it is deliberately provided for just a few generations of a frequently-upgraded software.
5. This potential obsolescence extends to file formats such as pdf, tiff, jpeg, and png commonly used by PCC staff. File formats closely linked to commercial programs are most vulnerable, but all file formats are constantly being developed and new versions made available.

4.2 PCC has a strong regime of back-ups, daily and longer term, and rules against storing data on independent media such as memory sticks, CDs or laptop hard drives. It is moving towards using the cloud to store all data - essentially buying space on other people's servers. It is, however, making no provision for keeping records accessible during and despite technical change apart from ad hoc arrangements such as those made during the recent move from Windows XP to Windows 7. The proposed move to Windows 10 brings threats to more programs and databases. Advertising for Preservica, a leading digital preservation software package, suggests that any record required for over 10 years needs protection from these risks. While that probably exaggerates the problem, the risks are genuine.

- 4.3 A second area of vulnerability is the human factor:
6. 20 years is a long time in someone's career. Staff leave, change responsibilities or retire. Those who remain may have little interest in records they created decades earlier; new staff are unlikely to learn about or value all the important records created by their predecessors.
 7. The structure of the council will change, leaving 'orphaned' records in no-one's charge.
 8. The council's functions are likely to change as they did at local government reorganisation in 1974 and 1997. New ones will be taken on, such as public health, and old ones surrendered although legal responsibility for keeping the records of these functions may remain.
 9. Records may be deleted by accident, because their significance was not known, or deliberately to conceal unpleasant truths, mistakes or fraud.
 10. If they are not controlled, records may be opened illegally and information accessed or amended.
 11. If access to them is not controlled and monitored their value as evidence will be compromised.
 12. If they are encrypted, passwords or keys will be forgotten or lost - this has recently been reported from Public Health and has also occurred in Archives.
 13. Deliberate cyber-attacks may be made on them.
- 4.4 Without a planned approach, long-term access to records will become inconsistent. Staff may not know that records on a particular subject exist, or if they do, may be unable to find them. It may not be clear who created particular records or why. Their reliability and legal admissibility will be questionable, since no record will have been kept about what has been done to each file in the course of its existence.

5 The business impact of digital preservation measures

- 5.1 While the council has the right records, it is able to carry out its functions effectively. Without them its ability to perform its functions and serve its citizens is seriously impaired while it and its officers may become liable to sanctions. The principal benefits of preserving access to digital records and costs of failing to do so can be summarised as follows:
1. Moral and reputational: PCC has a duty to serve its citizens effectively. Being unable to access its older records easily will render it less effective. Moreover public knowledge of the council's failure to keep records will undermine confidence in the council as a body and in individual councillors and staff, damaging their political standing, morale and career prospects.

The city's position in future local government reorganisation will be weakened.

2. Having access to older records and data allows them to be drawn on in the future. Decision-making will be better-informed and quicker, while collecting new data may be unnecessary. For example, effective management of contaminated land is reliant on recording past land use and condition, and drawing on the information to guide planning consent for new development.
3. Preserving digital records and managing them over even the short term using a cataloguing system can save unnecessary costs. The council's inability to find past drawings of electrical circuits in its buildings, for example, has meant expensive research and redrawing will be needed.
4. Failure to preserve electronic records and to be able to demonstrate that they have been stored secure from alteration will make it harder to defend the city's interests in litigation - such as the recent action over the contract for maintaining Portsmouth's roads in which the council brought to court 40 large boxes of paper records. Poor record-keeping contributed to the city's difficulties satisfying claims against it from residents of the former Cottage Homes, a type of issue which will easily recur if its digital records of children in care become inaccessible.
5. Loss of digital records could lead to fines for failing to keep records required by legislation or regulation, such as those outlined in the corporate retention schedule.
6. Not controlling access to records containing personal information over the longer term increases the risk of breaches of data protection legislation such as the General Data Protection Regulation that comes into effect in 2018. Fines for such breaches could be substantial.
7. Sustained or increased reliance on storing paper records because of lack of trust in electronic records would increase direct costs. The city currently has some 37,000 boxes of paper records in its Modern Records store; there are a further 7,000 (approx.) stored offsite at a cost of over 20p per box per month. The latter are predominantly children's social service records with lengthy retention periods. Accessing them incurs an extra cost for delivery.

6 Action being taken by other organisations

- 6.1 In the UK the lead on digital preservation has been taken by The National Archives and university computing departments. The University of Portsmouth, for example, has a 'Future Proof Computing Group' and runs postgraduate courses in digital preservation to produce people trained to work in this field. The National Archives early this century funded the creation of software to manage digital preservation of the government records in its

hands - this application is now marketed by its developers, Preservica plc. These organisations are concerned with research, education and cultural heritage. Their interest has been in promoting digital preservation within their own sectors and staff. An umbrella organisation, The Digital Preservation Coalition, <http://www.dpconline.org/>, established in 2002, has been spreading the digital preservation message more widely. It provides training and support to all sectors, and publishes the valuable Digital Preservation Handbook. Its membership includes British and Irish national libraries and archives, and many universities, but also NATO, Lloyds, HSBC, The Bank of England, the Grosvenor Estate, the BBC and the UN. Preservica customers include Oxfordshire & Suffolk county councils, Transport for London, the Met Office and HSBC.

- 6.2 Nevertheless few British local authorities have as yet taken active steps in digital preservation. No local authorities are members of the DPC, reflecting a lack of awareness within the sector.¹ A Records Management Society/Archives and Records Association survey in 2008 revealed that only 8 of 38 local authority archive services had then created functional digital archives and they were not managing the digital records of enduring value created by their authorities. This was perceived as a vicious circle - no records were being deposited because they had not created a digital repository and the repositories were not being created because few records were being deposited. A 2016 survey by the Information Governance Initiative (an American think-tank) of 196 organisations across the Western hemisphere showed only 11% had a 'standards-driven digital repository', while 44% were still considering what action to take and 30% had not begun to think about the issue.
- 6.3 Concerned local authority archives in our region have formed a group, Archives First, to explore options for digital preservation in their authorities. Besides Portsmouth's Library & Archive Service, it includes representatives of the following councils: Dorset, West Sussex, Surrey, Reading, Isle of Wight, East Sussex, Gloucestershire, Kent, Southampton, Hampshire and Wiltshire. None of these councils appear to have taken significant steps in digital preservation as yet, with the possible exceptions of Dorset and Gloucestershire. What has been done is limited to the historical archives. Preservica is offering this group a collaborative purchase of the company's services, which would cost less than an individual purchase: at a post-lunch conversation in October Preservica quoted the Library & Archive Service

¹ Partly perhaps because local authority records managers and archivists are mostly members of the Archives and Records Association and therefore access DPC resources through that body.

£10,000 pa for its application and storage of 2 terabytes of data, but perhaps as low as £5,000 if we go in as part of the Archives First group.

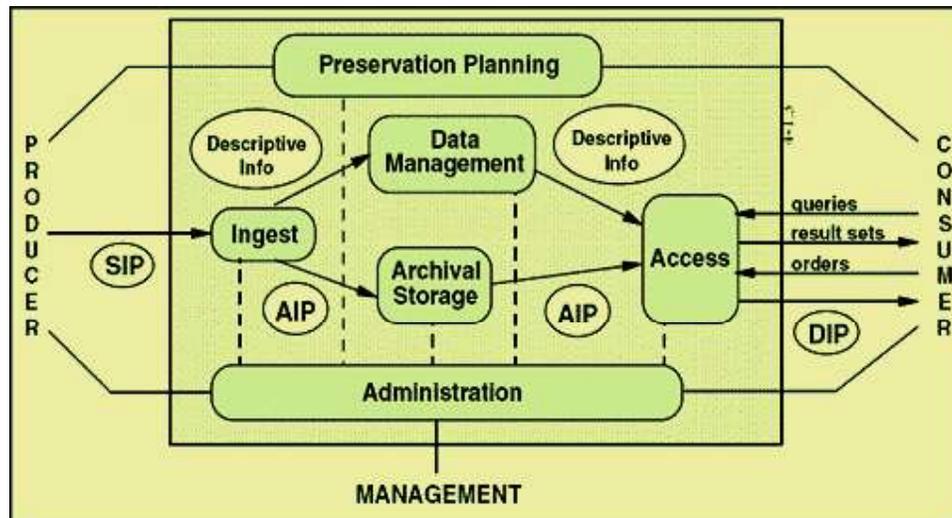
7 Potential courses of action

- 7.1 To ensure that its records remain usable for the required periods, the council needs to preserve the bitstream of each record, and provide an environment within which the file can be opened and read. It must also keep them secure from tampering and maintain intellectual control over them, so it knows what records it has, who created them and why.
- A. The first option is to carry on as now, doing nothing more than migrating records from one format to another and amending software to work in new operating environments as hardware and systems are upgraded.
 - B. A second option is to create a digital repository to protect and maintain access to its electronic records. Such a repository would be similar in concept to the Modern Records Service that secures the council's paper records and ensures disposal at the due date. A digital repository need not be a physical thing, requiring its own servers and space - it can be a virtual space within the council's existing servers or the cloud. There is an international standard for digital repositories to use as a guide: The Open Archives Information System Reference Model, ISO 14721:2012
- 7.2 Option A is the cheapest option in the short term. It requires no more than the current investment in IT staff; more resources can be added at pinch points like major upgrades. This approach makes no provision for intellectual control of the records (threats 6-12), nor will it provide evidence of how records have been treated to keep them accessible and thus give an audit trail proving their reliability as evidence. There are no safeguards against accidental damage to records and no way of checking that they remain undamaged until someone tries to open them - by when back-up copies may be corrupted, too. There is also no provision for checking that after migration into newer formats that records are still usable. It is certain, therefore, that in the long run the council will find it does not have records it requires. Recovering the lost records may be impossible or involve expensive 'digital archaeology'. In the long term, therefore, the costs are unquantifiable but likely to be significant.
- 7.3 Option B will be more effective, producing an archiving system which will be trusted by staff, external users and the courts. To create such a system will require
1. staff expertise to create it and promote it to 'depositors' within record-creating and using staff,
 2. staff time to manage it and execute the processes once it is in operation,

3. software, either drawing on open access software or buying an 'off-the-shelf' application, to carry out the digital preservation actions required and capture the appropriate metadata,
4. 'cataloguing' software to enable potential users to find the files when required, and
5. purchase of memory space to store the originals and necessary copies.

8 An OAIS digital repository in action

8.1 This diagram expresses the elements of the Open Archives Information System Reference Model:



SIP = Submission Information Package, AIP = Archival Information package, DIP = Dissemination Information Package, ie what is accepted into the digital archive, what is stored there after ingest actions and what is provided to users of the information.

8.2 In the digital archive the data is 'ingested' - taken into the archive, checked for viruses, its file format identified, copies made and stored in a separate physical location, and the bits counted to enable checks to be made for future damage to the file. The resulting metadata and information about the creator and content is stored and managed, along with the results of periodic checks on the file. If damage is discovered the file is replaced by one of the copies. When access to a given file format is likely to become difficult the file is migrated into an accessible format. Future users of the records access information about the records in the repository (or staff do so on their behalf) through a catalogue and the users are provided with copies of the files. The integrity and reliability of the preserved files is maintained; if required, disposal can be scheduled into the records' life in the repository.

- 8.3 Access to files in outmoded formats can be maintained in three ways:
1. by 'emulation' - creating on modern computers an environment which allows old software to run
 2. by the 'computer museum' approach - maintaining working examples of all types of computer and software versions needed to read the preserved files
 3. by 'migration' - transferring documents in formats which are no longer supported by PCC to formats that are, eg .doc to .docx, PDFa to PDFb.

Approaches 1 & 2 would certainly be beyond the resources of the council and poor value for money.

9 Recommendations

- 9.1 Failing to take steps to ensure the long-term preservation and security of its digital records will undoubtedly lead to problems for the council. The authors of this report, therefore, believe the council should adopt a Digital Records Preservation Policy and an accompanying strategy. The key element of this strategy would be to create a digital repository based on migrating file formats to maintain readability; this repository should follow the standards of the OAIS model rather than attempting to reinvent them.
- 9.2 We need to determine, however, the council's requirements: how to adapt the model to PCC's circumstances and organisation, and to gain some idea of costs. We suggest that the council set up a working group to examine the issues and in particular to consider the following:
1. What electronic records does the council wish to preserve for 10-15 years or longer? For example just records identified in the corporate retention schedule as having retention periods of 10 years or more, or also the corporate website and other ephemeral records?
 2. What is the volume of such records now and what is the annual creation rate of new ones in these categories?
 3. What is the range of file formats they are in?
 4. What metadata would PCC want to capture about records being preserved?
 5. To what extent would PCC want to automate the processes of digital preservation?
 6. What would be the most appropriate software to use to manage the repository? For example, fully open source, such as that created by Archivematica, or proprietary?
 7. If proprietary, how much would it cost and which application to choose?

8. Should PCC opt for storage in the cloud or on its own servers? And if in the cloud, fast- or slower-access storage? (IT is moving towards cloud-based applications, so the cloud is likely to be the preferred option.)
 9. What software should PCC use to maintain intellectual control of its records in the digital repository - (eg CALM, currently in use by the archive section, Spydus, currently in use by libraries and to which archives are planning to migrate, or some alternative?)
 10. Depending on which solution is adopted, what staff would be needed to manage the archive, to execute the processes and provide access to users?
 11. Could this be done within the existing Modern Records Service?
 12. Should the repository be managed within IT (which has just given up information management), Libraries & Archives (which manages the Modern Records Service), or some other department?
 13. Would PCC want to provide access to catalogues of the records to staff outside the digital repository and to the public?
 14. Should PCC join the Digital Preservation Coalition <http://www.dpconline.org/> (an organisation providing advice, resources and training - the city's archivists have access through their membership of the Archives and Records Association, already part of the DPC)?
- 9.3 The group would draft the Digital Records Preservation Policy. Given the technical nature of many of these questions and present staff expertise in records management and digital preservation, the group should include IT and archive staff, perhaps supported by someone from Information Governance.