03 Working with record creators

Introduction

This chapter of the Workbook provides a broad overview of the workflow which takes place during the first two phases in the lifecycle of a digital archive (see p. 2), from record creation through to transfer into archival custody. Much of it is based on Paradigm experience, but many of the procedures developed by the project may be more generally applicable to the kind of digital and hybrid personal archives which will increasingly be acquired by collecting institutions.

By way of initial preparatory work, the project team formulated some ideas about the nature of personal archives and the ways in which the digital environment might change their shape and nature. More detailed research was carried out on the structure of existing personal archives of politicians held at both the Bodleian Library and John Rylands University Library.

Another of the project team's earliest tasks was to produce a deposit agreement for use with participating politicians. Whilst this is specific to the project, the team explored issues raised by the digital environment in relation to gift and deposit agreements more generally, and this resulted in a model gift agreement (see *Appendix A: Model gift agreement*) which might form a template for adaptation and use by other collecting institutions.

The team's investigation into the personal archives of politicians helped to inform the preparation of a records survey for use with the politicians participating in the project. The survey forms a crucial initial stage in the acquisition process and provides important information for later stages of work, such as the transfer process, appraisal and creation of metadata. The transfer protocol developed by the project is also outlined here, although it is acknowledged that this is only one of many ways in which digital archives might be transferred from creator to repository, and the method adopted may depend on the approach to collection development (see Chapter 02 *Collection development*) being taken.

Many of the procedures developed during the course of the project will already be familiar to archivists: sound traditional archival practice provides the foundation of Paradigm's acquisition and accessioning process, and this has simply been adapted to take into account the particular challenges raised by the digital environment. The procedures outlined here can also be mapped to the functions set out in the OAIS Model,in order to demonstrate that relevant OAIS mandatory responsibilities are being fulfilled, primarily the obligation to:

- Negotiate for and accept appropriate information from information Producers.
- Obtain sufficient control of the information provided to the level needed to ensure Long-Term Preservation.

The nature of personal digital archives

Defining 'personal archives'

During Paradigm's initial visits to politicians and their staff it became obvious that in order to explain the project brief and the significance of personal political archives a clear definition of what is meant by 'personal archives' was needed. Pinning down a clear definition of personal archives (also known as private papers, personal records and manuscripts) and why they are important is challenging. Individuals are complex, multi-dimensional entities and throughout their working and personal day-to-day lives they adopt various roles, the boundaries of which are often blurred.

Personal archives will often comprise personal, occupational, professional and special interest records. The politicians participating in the Paradigm project were naturally reluctant to place current personal material, or confidential records relating to third parties, in a library, especially for the purposes of a testbed project. Therefore, while the project explored the impact of digital living on the more personal types of personal archival materials, the exemplar material accessioned during the course of the project reflected the professional, rather than the personal, roles of participating politicians. As such, it included policy briefings, drafts of speeches and other internal documents generated by the politicians' private offices.

A special relationship: why curators should talk to creators of personal digital material

The information found in personal archives gives readers a unique, human perspective into historical events. The wealth of correspondence and papers produced by actors and observers provide details which the official record cannot. Letters, diaries, collections of press cuttings, photographs and drafts which were not created for public consumption are amongst the most valuable historical sources. These sources are increasingly created in a digital format and because of this many records have already been lost, or are in danger of being lost. Personal archives are especially vulnerable because they sit outside organisational structures which could provide for their maintenance by imposing standards or policies relating to digital record keeping.

Record creators are not always aware that even the short-term survival of their digital material is at risk due to media degradation, corruption or accidental loss. Most individuals only have limited support from IT professionals; any technical support they receive tends to be informal advice from IT-literate friends or acquaintances. Often record creators have little understanding of the technology which underlies and facilitates their digital record keeping, and no great interest in delving into its complexities. They may also be unaware of the potential significance of their digital records and manage them in a haphazard or inconsistent fashion, deleting historically valuable records when they have outlived current needs.

By working with creators of historically significant personal archives curators can increase the rate of survival of these valuable resources. Even in the early stages of the Paradigm project it was possible to discern changes in the record keeping behaviour of politicians' staff working with Paradigm archivists. The staff of one politician decided to place weekly diary briefings in a folder rather than deleting them once their immediate usefulness had expired. Several participants remarked that they had never thought of the records they dealt with as having any long-term historical interest, but on reflection they appreciated the potential value of the material to historians and other researchers in the future. Such reflection is likely to lead to changes in record keeping behaviour and this can be extremely helpful for archivists – making the task of appraisal (see Chapter 04 *Appraisal and disposal*) much easier and resulting in a more complete historical record.

it is also worth noting that many of the paper records generated by politicians are at risk. Countless records without administrative value are routinely destroyed during the parliamentary recess. After a general election campaign, a change of brief, or the redrawing of a constituency boundary, destruction can be even more extensive. Other domains may be vulnerable to different changes, for example constant change in domains where project working is the norm. Once paper records reach the archive, however, their preservation is largely a passive exercise; management decisions regarding appropriate physical storage and access conditions can be applied in blanket fashion.

Maintaining contact with record creators over time is therefore of the utmost importance for archivists collecting personal digital archives. It may take years for record creators to develop enough trust in the repository to consider depositing their most sensitive or confidential papers. This obviously requires a commitment of staff time and resources on the part of the collecting institution; maintaining contact over a long period isn't easy, especially in the case of politicians' offices, where there is often a high turnover of staff.

Why is it important to preserve digital records digitally?

Where personal digital archives exist they should be preserved digitally, to retain their inherent qualities and relationships. While some page-oriented records lend themselves to printing, many would lose salient characteristics if transferred to paper. These important characteristics include formatting, relationships, intertextuality and other functionality.

Preserving records digitally will also allow archivists to preserve the 'original order' of personal archives, whilst also providing a functionality that enables users to rearrange the material by date, author, or other property. Full-text searching is another possibility which interests many researchers.

Preparatory work: investigating the structure of the private archives of politicians

From the outset the Paradigm team were aware of the distinctive nature of personal archives and before addressing the challenges of digital preservation the project archivists felt it necessary to re-examine the nature and structure of private archives. The institutional experience of the Bodleian and the JRUL assisted the project in this matter; both libraries have collected the papers of politicians and other prominent individuals for many years.

Comparing existing holdings with records created by today's politicians

Naturally, Paradigm began by exploring existing holdings, identifying the record types found in these and thinking about how emerging technologies might have changed the creation, format and survival of such political records.

Looking at these records also helped to establish what roles, activities and relationships personal archives bear witness to, as these are important criteria when selecting records to preserve and appraising collections: functional appraisal (see p. 40), for example, involves looking at the functions carried out by the record creator, identifying the functions that merit documenting and selecting the records which best represent those functions, rather than starting with the records themselves.

As the project continued, experiences of working with contemporary record creators also contributed to Paradigm's understanding of personal archives.

The types of records traditionally found within the papers of an average politician include constituency and parliamentary correspondence, engagement diaries, political briefings, speeches, press releases, photographs, election material and personal correspondence. In a digital age the same series still exist, although technological advances may have changed the actual manner and format of their creation, use and storage.

Correspondence

Correspondence, personal and professional, is a key component of most personal archives. New methods of communication have resulted in the decline of letter-writing, but it is possible that mediums such as email and instant messaging could capture many exchanges which once took place via phone as well as by letter. Unsurprisingly, email is now a primary communication tool for many politicians and some of the private offices the project has worked with receive an average of 200 emails a day. One office has an email archive dating from 2003 which contains an overwhelming 70,000 emails. Paradigm's Academic Advisory Board considers email to be the digital record-type of most historical value, but researchers will need sophisticated navigation and discovery tools to manipulate an archive on this scale.

Engagement diaries

Engagement diaries, which contain so much valuable information, are now likely to be in the form of a Personal Digital Assistant (PDA), or even a MS Word file.

Documents and drafts
These include many kinds of papers, such as political briefings, speeches, press releases, internal publications, etc. Many personal archives at the Bodleian and the JRUL contain drafts of such papers as well as the final versions, although how many drafts individuals create and whether they keep them varies from one archive to another. Historians and curators are concerned that the use of word-processing software has resulted in the decline of the draft, which previously enabled researchers to trace the development of the writer's thoughts. Although it is possible to save versions of a digital document during the drafting process, many authors simply overwrite earlier versions or delete records of the drafting process.

Digital photographs
The advent of the digital camera has led to the accumulation of large quantities of digital images taken by individuals, many of which will have no analogue equivalent. Paradigm's experience is that most of these photographs have precious little contextual information, although contact with the creators and cross-referencing with engagement diaries could assist cataloguers in creating meaningful descriptive metadata. This is also true of photographs found in the traditional paper archives of politicians held by the Bodleian and JRUL which often lack basic descriptive information and require the cataloguer to undertake some detective work to place them in context. Some digital images may at least provide a definitive date of creation.

Personal websites and weblogs
Formats, such as websites and weblogs, do not have an obvious historical equivalent, but their role in politics is increasing dramatically. The 2005 general election saw many new candidates, as well as those defending seats, use a weblog as part of their campaign strategy. A growing number of politicians use blogs as a means of engaging with the electorate on day-to-day issues. The blog - a frequent, chronological publication of personal thoughts and web links - is a curious mix of newspaper column, soapbox, diary entry and index to external items of interest. Such topical personal material is likely to have long-term historical interest and ranks highly as valuable historical material with the project's Academic Advisory Board. Institutions collecting personal archives should consider the websites and weblogs of those individuals as a component of their personal archives and seek to preserve them.

Digital audio and video
Archivists at specialist sound archives have preserved audio recordings of speeches for a number of years. It is likely that non-specialist repositories will increasingly accession personal collections holding significant amounts of audio recordings due to the explosion of new technologies and gadgetry such as the iPod and other MP3 players. Personal digital video is also likely to grow alongside the increasing popularity of web services for video material.

Election materials and press releases
Relevant materials are created in all the formats above. Election material, along with press releases, have been created for the public domain and they are examples of personal political records that can be opened shortly after they have been accessioned.

Newscuttings and scrapbooks
Many archives contain newscuttings and offprints. The digital equivalent of these could include the history of an individual's web-browsing, or a list of the websites bookmarked and RSS feeds subscribed to.

Factors affecting the depth and breadth of personal archives

Various factors can affect how comprehensively an archive documents all aspects of its creator's life. Concern about privacy and confidentiality restricted the scope of material deposited as part of the Paradigm project. Another pertinent issue, particularly during an election year when much of the material was being accessioned, is a politician's lack of time for preserving personal archives. However, during the course of the project's work, it became clear that there are a number of factors which can affect the survival and shape of personal archives.

Individuality

Examples from Paradigm's digital and paper accessions show that some personal archives document their 'creators' more comprehensively than others. Despite certain similarities between Paradigm's participants, in terms of role and routine, the 'personal' element, or 'individuality' of personal papers should not be forgotten. As Sue McKemmish expounds in 'Evidence of Me', her seminal paper on personal record keeping, individuals exhibit a broad range of 'personal record keeping behaviours': some individuals keep every letter they ever receive, others keep nothing and rely solely on memory. The vast majority of people fall somewhere in-between these two extremes. The heterogeneity of people, records and systems means that Paradigm's dealings with each participant have varied to some degree. The procedures adopted by curators of personal digital archives will need to be sufficiently flexible to take account of this.

Selection for preservation by the creator

It is not only what is routinely saved in the course of business that influences the extent and individuality of a person's records, it is also what is volunteered for long-term preservation. By way of example, the office of one Paradigm project participant permitted the entire contents of an email inbox to be copied and archived (aside from a few confidential files); this is in marked contrast to other participants for whom email was definitely off-limits. It is the very recentness of these records and the fact that emails and other digital records are easily forwarded, which makes office staff nervous, particularly in the unforgiving world of politics.

Third parties in personal archives

Political 'personal' archives like those accessioned by Paradigm are distinctive in that they are not generally the work of one person. Rather they are a joint enterprise with much of the day-to-day correspondence and office papers being created by the MP's Personal Assistant, or the constituency office staff. A large amount of the material is also generated by individuals and organisations outside the employ of the politician. This includes correspondence from constituents, colleagues, special interest groups and lobbyists, as well as briefings, research papers, etc., circulated by the central political party office. This means that there are third party interests besides the politicians' to consider, such as privacy and intellectual property rights, and this is likely to be the case with most personal digital archives.

Many Paradigm accessions were contemporary records and were inevitably likely to contain more personal data about living individuals as defined by the *Data Protection Act (1998)* than archival materials accessioned much later after creation (see p. 250). Politicians were made aware of the exemption in the Act which covers the activities of archives and libraries in this area (s. 33), but some participants raised ethical concerns that the creators of the records might not agree to the placing of their email in an archive. This is interesting as 'paper' letters from third parties are captured in private archives held at archives and libraries the world over. Both the recent date and digital format of the material acquired by Paradigm increases the sensitivity of this issue.

For a politician of the governing party, matters are further complicated by the potential for overlap between the content preserved in a personal archive and that contained in public records preserved under legislative requirements by The National Archives (see p. 247). In such instances, the curator of the personal archive would need to refer to The National Archives regarding the classification of similar material. Any public records included in a politician's personal archive would not be acquired as part of the archive.

Technological constraints

The types of digital records that are created and filed for medium to long term preservation are also dependent on the computer expertise and technical know how of politicians and their key office staff. For example, some offices are familiar with presentation software, PDF creation, website development, desktop publishing and digital cameras; others are less comfortable with emerging technology and use their computers for little more than email and word processing. Another factor impacting on the extent of material saved is the quantity of storage space available. Emails are particularly vulnerable to deletion due to inbox quotas.

Are politicians typical of other creators of personal archives?

Paradigm's investigation into the structure of personal archives focused upon politicians and some of its findings will not be relevant to those dealing with the personal archives of other types of in-

dividual. Indeed, perhaps politicians are not the most typical creators of personal archives as it is obvious that they do not personally create many of the records which fall under their name. However certain aspects of personal record keeping are universal, namely the sheer variety of record keeping behaviours and the idiosyncrasy of the individual. *The Guidelines for creators of personal archives* (see *Appendix B*) have been written with all types of individuals in mind and focus on easily implemented practical solutions. It is hoped that these can provide a basis from which digital curators might produce refined guidelines targeted at individuals working in specific disciplines (e.g. writers or scientists).

Donation and deposit agreements

Introduction

Drafting an appropriate deposit agreement was one of the earliest tasks Paradigm undertook. It proved impossible to find other examples of deposit agreements drafted for personal digital archives, so the project team began by looking at the kind of donation and deposit agreements in use for traditional archive material and thinking about how these should be adapted and augmented to meet all the requirements of the digital environment.

The project deposit agreement was produced for a very specific purpose; participating politicians agreed to loan their digital archives to the Bodleian and JRUL on the basis that the content was treated as confidential and access was limited to project staff. However, many of the concerns raised by the Paradigm participants are pertinent to personal digital archives more generally, particularly the issues of: privacy, confidentiality and data protection; access by third parties, especially by means of freedom of information requests; and IPR (Intellectual Property Rights) in the content of the archive.

Paradigm therefore explored some of the major issues which should be covered in any donation or deposit document dealing with hybrid or digital personal archives, with a view to drawing up a model agreement.

Issues which should be covered in a donation or deposit agreement for personal digital or hybrid archives

Any agreement needs to be legally sound, yet easily understandable by donors or depositors. It should clearly set out the obligations of both donor/depositor and archive repository. The content and extent of the archive should be set out in the schedule to the agreement, indicating (in the case of a hybrid archive) the relative proportions of hard copy and digital material. Different conditions may apply to hard copy and digital material (e.g. in relation to access by third parties) and where this is the case it should be clearly stated in the body of the agreement.

Some of the issues which should be covered in an agreement include:

- Establishing unequivocally the current ownership of the archive itself.
- IPR: the donor or depositor of the collection may be a primary copyright holder in its content, but there are also likely to be many third-party copyright holders represented in the archive. The donor or depositor should be asked to clarify conditions relating to the material in which they hold copyright, both in relation to:
    - Preservation: the repository should seek explicit permission from the primary copyright holder to undertake preservation actions on the digital component of the archive; these can range from simple backup procedures to format migration and involve making multiple copies.

- Access: the copyright holder may be offered a range of options. For example, they may choose to: grant licence to the repository to carry out certain actions (such as making copies for researchers in accordance with fair dealing regulations, granting permission for the publication of short quotes, or making the copyright material accessible remotely rather than limiting access to researchers in the reading room); or request that all requests for copies (other than those for non-commercial research) be referred to them for permission. They may also choose to transfer copyright into the ownership of the repository, which would then be responsible for all decisions relating to the material.

The greater proportion of material in any archive is likely to be in the copyright of multiple other individuals. The repository can only grant access to this and provide copies of it within the stipulations of copyright legislation and fair dealing regulations; an undertaking to act within the boundaries of copyright law should be given in the agreement. Currently UK law is ambiguous about making multiple copies for preservation purposes where permission has not been sought, but the Gowers Review (see p. 263) may change this situation. See the Chapter 09 *Legal issues* (p. 252) for more information on IPR matters.

- Data protection and confidentiality: the donor may wish to place closures on confidential or sensitive material in the archive, to ensure that it is not made available to third parties for a stipulated period of time. It is useful to record both closure periods and reasons for closure in order to assist the repository in dealing with future freedom of information (see p. 265) requests (although it should be noted that deposited collections are not necessarily subject to the *Freedom of Information Act*). It should also be made clear that whilst libraries have certain exemptions from the *Data Protection Act* (see p. 250), all personal and sensitive personal data about living individuals represented in the archive will be treated in accordance with the Act, and in particular will not be used in any way which might support decisions about or cause substantial distress to individuals. This may mean imposing lengthy closure periods on some material.
- Other access-related issues: the agreement should make it clear that the repository is acquiring the archive in order to make it accessible to researchers in the future and establish the conditions for this (lengthy closure periods will usually apply). There should be a clause establishing the right of the repository to claim legitimate exemptions under the *Freedom of Information Act* while material is closed to readers.
- Exclusive ownership: because accessioning a digital archive may involve taking copies of records rather than the 'originals', it is crucial to ensure that the donor/depositor grants the collecting institution the right to hold (whether by gift or loan) the sole complete research copy of each digital record in their archive. This is particularly pertinent in the case of archives which are likely to be of high monetary value (e.g. the papers of a well-known writer).
- Appraisal: the agreement should establish the right of the repository to appraise and weed material in the archive; the donor/depositor should be given the right to decide what happens to weeded records. See Chapter 04 *Appraisal and disposal* for more information.
- The agreement should make it clear that the collecting institution will catalogue the archive and create the (often extensive) metadata required for long-term preservation; copyright in catalogues and other metadata will be held by the institution, as distinct from the copyright in the archive material itself.
- The collecting institution should give an undertaking to look after the archive material responsibly and securely (in perpetuity, in the case of a gift).
- If outsourcing the storage or processing of digital materials to a third party, permission should be sought for this from the donor/depositor.

- The schedule to the agreement may establish how accruals to the material will be managed and the frequency of any future accessions. The detail will depend on the selected approach to collection development (see Chapter 02 *Collection development*).

Deposit and gift agreements produced by the Paradigm Project

In order to simplify the administration and management of its holdings, ideally an archive repository should base all its donation and deposit agreements on a single template which is uniform within the institution. However, in reality some flexibility is likely to be necessary, as is the case with traditional archives, and agreements will often have to be tailored to a specific individual or situation.

Paradigm produced a model gift agreement for use with digital and hybrid personal archives (see *Appendix A: Model gift agreement*). This might form a basic template for other collecting institutions to build on and adapt to their own requirements. Please consult with your legal team to ensure that the agreement is appropriate to your context.

Paradigm also developed a short-term Deposit Agreement which was specifically designed for use with the politicians participating in the project. For the purposes of the project, digital archives were deposited on the understanding that their content was confidential and that the archives would be closed to third parties; access was restricted to project staff and individuals accredited by the depositor.

Surveying collections

Introduction

An important first step when weighing up the potential acquisition of any archive is to establish a broad overview of the archive's content and context. This is particularly important in relation to digital and hybrid personal archives, and Paradigm's investigation into the structure of politicians' papers informed the production of a records survey for this purpose.

Records surveys are traditionally seen as the preserve of institutional records managers, who have some degree of control over the creation and management of records during their active lifetime. In an environment of hybrid record keeping these surveys typically seek to gather information on: the types of record series maintained in an organisation or department; the functions that give rise to these record series; the creators and users of the records; their frequency of use; the length of time for which the records are required for specific purposes; responsibility for maintaining the records; the rate of accruals; legislative requirements for keeping the records; access and security issues; confidentiality and sensitivity of content; and the extent and format of both hard copy and digital components of the record keeping system. The results of records surveys are used to gain control of records, bring them into a managed environment, identify active, semi-current and inactive records, and establish filing schemes and retention schedules.

The digital environment requires archivists working in collecting institutions to take a much more pro-active approach towards collecting and to become involved with donors and depositors at a much earlier stage in the records lifecycle. Records surveys are therefore likely to form an important tool of the digital archivist or curator, and may even influence the record keeping practices of archive creators – especially if the repository chooses to take the post-custodial or snapshot accessions approaches to collecting (see pp. 11-12), which involve fostering an ongoing relationship between repository and donor/depositor over a period of time.

In the records management field, surveys are either carried out by physical observation and examination, or by questionnaire. In the world of personal record keeping, a combination of both approaches is probably most effective. The Paradigm archivists therefore sent their survey questionnaire out in advance, in order to prompt the project participants to start thinking about their record keeping practices. This was followed up by a visit from one of the project archivists, during which the detailed survey was carried out. On-site visits enable the record creators to ask any questions arising from the questionnaire; they also afford the archivist an opportunity to augment the survey information by interviewing the record creator(s), obtaining job or role descriptions (where records are created in an organised context like a politician's constituency office) and any existing file schemes or indexes which are maintained, as well as any other contextual information which might enhance understanding of the archive.

Paradigm also explored some other useful ways of gathering information about the digital components of a personal archive for survey purposes – principally by creating screenshots and capturing directory structures; both of these methods can provide important information about the original order and record keeping practices of the donor/depositor. They also help the archivist to identify records series of historical significance. Paradigm tested a number of tools for capturing directory structures (two are explored in Chapter 04 *Appraisal and disposal*, see p. 45), which provide a useful means of extracting information about an aggregation of folders and files. The project also explored the potential of tools developed for the digital forensics community for the secure acquisition and investigation of digital materials. There are several open source and proprietary tools available that merit more detailed investigation by the archival community.

The records survey

The Paradigm records survey (see *Appendix C*), sent out to politicians' office staff, probably bears a closer resemblance to a typical institutional records survey than the kind of questionnaire that might be produced for more obviously 'personal' archives (which are largely the product of a single individual). However, it could be adapted for different types of archive creator. It seeks information on: specific records series which are likely to be of historical interest; the content of the records; the structure of the archive; the functions and roles of office staff and record creators; rights and responsibilities in relation to the records; and the technical environment in which record creators are working.

How does the records survey assist the digital archivist?

Much of the information provided in the records survey is valuable for the purpose of administering and managing the archive once it reaches the repository, and for the creation of preservation and descriptive metadata. Specific ways in which a records survey can be useful for archivists responsible for personal digital archives include the following:

- The survey allows a degree of pre-acquisition appraisal (see p. 37) to take place: the survey should provide sufficient information for the archivist to make appraisal decisions based on some of the key characteristics of the records (see p. 35). In the case of records which are obviously of little historical value, records which are also preserved elsewhere (e.g. at The National Archives, in the case of some politicians' papers), or records which obviously fall outside the archive repository's collecting policy, the archivist might decide not to select these for accession. The survey also allows the archivist to determine where duplication occurs – particularly between the hard copy and digital components of an archive – and to make decisions on which version should be preserved as the 'master' archival copy of a record. Technical information on the file formats used by the record creator is also useful for appraisal purposes: if obscure formats are used, or if files are likely to be difficult to extract from their current technical environment, the archivist may decide not to acquire them as part of the archive.
- It prompts the archivist to ask about a range of records that may not be automatically volunteered by the creator, such as personal digital materials that are stored online.

- It prompts record creators to think about the significance of their personal archives and perhaps to take some responsibility for actively managing their own records; this may result in a more 'rounded' archive which comprehensively documents all the various roles of its creator.
- The survey establishes the identity of the principal record creators represented in the archive and the copyright status of the material; this information is crucial in enabling the digital archivist to manage the material, carry out preservation actions and make the archive accessible to third parties. This information will ultimately form part of the IPR metadata (see p. 141) associated with an archive.
- It also clarifies where material in the archive is likely to be subject to other legislation; the *Data Protection Act* (see p. 250) in particular is likely to apply to much of the material in a digital archive which will inevitably include data about living individuals; this is even more pertinent where records are acquired during the creator's lifetime (as in the regular snapshot accessions approach to collection development, see p. 11) rather than at the end of the creator's life or after their death.
- Information gained from the survey about how records are organised by their creator(s), and how the digital and paper components of an archive relate to each other, will inform the ultimate arrangement of the archive at cataloguing stage (see p. 176). Much of the administrative, biographical and provenance information required for the catalogue can also be obtained during the survey process.
- Questions about hardware and software used by the record creator(s) are useful for immediate practical reasons: they establish what technology is available on site (e.g. are office PCs networked? Are there USB ports or CD writers?), what the accessioning archivist may have to learn before transferring records and what equipment they may have to take with them to undertake the transfer process. If unusual or obsolete file formats are identified in the survey, the archivist may need to acquire or seek out special skills, software or hardware before the transfer process can take place.

The archivist can augment all of this survey information through interviews and onsite visits to donors/depositors.

Creating screenshots

When surveying digital archives at the premises of donors/depositors it can be helpful for the archivist to create screenshots which can be saved or printed. These can be created using nothing more than the computer of the creator or with more sophisticated screenshot tools. Screenshots can also be used to capture system information about the creator's computing environment and to provide a record of the original directory structures used to manage email and other documents. This allows the archivist to identify important record series and folders, and to gain an insight into the creator's record keeping practices and the original order of the records. Screenshots are also useful in the transfer and accessioning process: they provide a visual aid for the donor/depositor, and allow the archivist and record creator to come to a clear agreement about the records to be selected for accession.[1]

Screenshot tools
Screenshots via the Print Screen key are limited. The image is constrained to the size of the screen, which may not include the whole directory tree, and it is not possible to target specific windows or regions of the screen. More sophisticated screenshot tools that can capture specific windows and scrolling windows are available and could form a useful part of the archivist's survey/accessioning toolkit.

1 Tips on how to create screenshots can be found in the online Paradigm Workbook. URL: <http://www.paradigm.ac.uk/workbook/record-creators/screenshots.html>

Capturing directory structures

Capturing the directory structure of an archive creates a record of the original order of digital materials accessioned by the repository, which can be used for the same purposes as screenshots; however, this has the added advantage of generating a single text file containing all the directory information, which can then be searched.[1]

Accessioning digital and hybrid personal archives

Introduction

The term 'accessioning' in the digital environment essentially denotes the activities which take place after the records survey and prior to ingest into the digital repository, i.e. 'the process of transferring the selected records and the records survey metadata (the Submission Information Package (SIP)) from the creator's computing environment to the accessions archive where the Archival Information Package (AIP) is generated; this represents phase 2 in the lifecycle of a digital archive (see p. 2). This process must take into account the concerns of all parties with a stake in the archive:

- The record creator will be concerned about the security of their data during transit. In the case of a busy politicians' office, creators will also be keen that the transfer process can be carried out quickly and with a minimum of disruption.
- The digital archivist is concerned to ensure the authenticity and integrity of the records throughout the process, i.e. that they have not been unintentionally altered in any way and that their original directory structures are maintained intact; this also needs to be documented fully in an audit trail.
- Authenticity will also be a prime concern of future researchers, who will want to be sure that the records they access are identical in all essential respects to those which left the creator's computer.

Paradigm developed a transfer protocol for the purposes of the project which attempts to take all these issues into account. However, it should be noted that this only represents one possible approach to the transfer process (transfer via removable media), and ultimately the process must accommodate a range of different methods which will vary according to the selected approach to collection development (see Chapter 02 *Collection development*), e.g. electronic transfers from the donor/depositor via a secure upload mechanism, or transfer via retired hardware and media when records have reached the end of their active life.

This section outlines the Paradigm transfer protocol and documentation, and points to practical 'how-to' guides for accessioning two commonly-encountered types of digital record. The paper-based component of hybrid archives is not covered because most collecting institutions will have well-established procedures in place for transferring traditional archives to the repository.

Transfer protocol

The Paradigm transfer protocol was designed to balance the needs of archivists with those of record creators, and to be both effective and relatively straightforward to carry out.

The goal of the protocol is to enable authentic records to be securely transferred from the premises where they are currently accessed or stored to the premises of the Library's Digital Archive. It in-

1 Tips on how to capture directory structures can be found in the online Paradigm Workbook. URL: <http://www.paradigm.ac.uk/workbook/record-creators/capturing-directory-structures.html>

cludes measures designed to:

- Preserve as much of the records' original order as possible.
- Protect the integrity of the records.
- Secure the records from unauthorised access.

The transfer protocol developed for the project is based on copying the records selected for accession to removable media.

Pre-accession assessment

The pre-accession assessment of the archive will normally include a site visit to conduct the records survey and some discussion with the depositor (or representative) via phone or email. The records survey has been developed to enable the digital archivist to gather the information required to undertake an accession which includes digital materials; it assists the archivist in:

- Discovering which hardware(s) and software(s) are being used to create records.
- Learning about any username/passwords which might be required to access or copy materials.
- Discovering which hardware(s) and software(s) might be used in the transfer process.

The transfer list

A transfer list was produced (see *Appendix D: Transfer list*) to document the transfer process; this forms part of the audit trail for each accession and supports the authenticity of its component digital objects. A separate list can be used for each directory structure, e.g. email directory; 'my documents' folder. It allows the archivist to record:

- The name and contact details of the owner of the archive material and the archivist carrying out the transfer. Both parties can sign and date the list.
- A reference to the deposit agreement (see p. 22), which sets out the terms and conditions of the transfer.
- A reference number for the piece of removable media being used, e.g. USB-1.
- The checksum value generated for the material. The checksum calculation is repeated when the records arrive at the accessions archive to ensure that they have not been altered in any way during the transfer process. Paradigm used the MD5 algorithm for checksumming at accession; the checksumming algorithm defines how the checksum is generated and several algorithms exist.
- The extent of the material, ideally in bytes (the size measurement required for preservation metadata purposes).
- A technical description of the material, including information on file formats, any passwords or encryption applied by the creator, and details of the PC from which the material was taken.
- A description of the material's content, including information on: directory and folder names; the identity of the principal creator; record types included; broad subject areas covered; and approximate covering dates.
- Restrictions which apply to the material, including an indication of: records which might contain confidential or personal data; records on which the creator wishes to impose access restrictions; and records which both parties have agreed will not form part of the accession (these should be securely deleted if they have been picked up during the copying process).

It is useful to supplement the transfer list with printed screenshots of the system information and directory structures as a visual aid. A copy of the list is offered to the creator and a copy remains with the archivist to form part of the audit trail.

Transferring records to removable media

The first site visit and survey should have helped the archivist to determine which records are to be captured, how they will be captured and how long the transfer process will take. This should be agreed with the creator and the archivist should now be prepared to make the first accession; for this they will need:

- A laptop with checksum software, encryption software, anti-virus software, CD-ROM reader/USB port.
- Removable media on which to store the records. The type of media should be agreed in advance with the depositor and will depend on the hardware available and the quantity of records to be captured. It might be a CD-ROM or USB stick, although in light of concerns about the security of data expressed by project participants, the Paradigm archivists ultimately opted to use a biometric protected USB-powered external hard-disk. It can be useful to take a selection of different media.
- A copy of the deposit agreement.
- A supply of transfer lists.

At the premises where the records are to be copied, the archivist should follow these steps:

- Where possible, records should be compressed in existing directory structures using lossless compression, such as a tar or zip file. This helps to maintain original directory structures, speeds up transfer and requires less storage space.
- These compressed files should be transferred to the removable medium.
- The removable medium should be inserted into the digital archivist's laptop and, using the MD5 checksum software pre-installed on the laptop, a checksum of each compressed file should be generated and recorded on the transfer form. An authenticity check is carried out when the records reach the accessions archive as part of a general 'health check' on the material, by repeating the checksum calculation and comparing the result with the checksum(s) recorded at the transfer stage.
- The transfer form should be completed and signed by the digital archivist and the depositor or their representative.

If the archivist is able to use a USB mass storage device, then it is possible to add a toolkit of compression, checksumming, screencapture and encryption software to this device and to run these while the device is connected to the creator's computer, rather than creating checksums and encrypting files using the archivist's laptop. Archivists should also explore digital forensics tools, which are designed for investigators to capture digital material that can subsequently be used as evidence onsite, as these have interfaces for the extraction of materials and can automatically create checksums for each file.

Accessioning common types of personal digital record

It is vital that the archivist is prepared for the first accession visit and arrives armed with sufficient knowledge to deal with the various record types identified in the survey, using the available technology (also identified by the survey).

Carrying out a digital accession is no easy task; even for the IT literate, orienting oneself on someone else's computer can be challenging. Paradigm believes that archivists need 'how-to' guides to assist them in acquiring some kinds of digital record that must be 'extracted' from their environments before their files can be transferred to a digital archive repository. Important examples include:

- Records, such as email and images, stored in various web-based services.
- Email accessed via a client.

Guides for extracting materials from web services must be regularly updated and guides for extracting materials via clients should reference particular versions of the software. Archivists will also need how-to guides for materials where the extraction process takes place at the digital archive repository:

- Extraction and migration of records stored on obsolete media.
- Extraction and migration of records stored in obsolete formats.
- Remote harvesting for personal websites.

Sample how-to guides

How-to guides for extracting two commonly-encountered types of personal digital record are available in the online Paradigm Workbook: harvesting websites using Adobe Acrobat Professional 7.0 and exporting email from Microsoft Outlook email clients.[1]

Useful resources

7-Zip File Archiver
Open source compression tool.
URL: <http://www.7-zip.org>

AccessData
Provider of digital forensics products and training.
URL: <http://www.accessdata.com/>

DirPrinting
A tool for viewing directory structures, saving them to a file and printing them.
URL: <http://www.majusoft.de/DirPrinting/index_en.htm>

Guidance software
Provider of digital forensic software and training.
URL: <http://www.guidancesoftware.com/>

Helix
Open source digital forensics tools on a bootable live CD.
URL: <http://www.e-fense.com/helix/>

1 Tips on Harvesting websites with Adobe Acrobat Professional 7.0 can be found in the online Paradigm Workbook. URL: <http://www.paradigm.ac.uk/workbook/accessioning/adobe.html>
Tips on Exporting email from Microsoft Outlook email clients can be found in the online Paradigm Workbook. URL: <http://www.paradigm.ac.uk/workbook/accessioning/microsoft-outlook.html>

Karen's directory Printer
A more sophisticated tool for viewing directory structures and file attributes, and saving them to a file or printing them.
URL: <http://www.karenware.com/powertools/ptdirprn.asp>

Microsoft – Help and Support
Provides some useful information about extracting data from Microsoft products.
URL: <http://support.microsoft.com/>

Open Source Digital Forensics
Reference for the use of open source software in digital forensics.
URL: <http://www.opensourceforensics.org/>

Sleuthkit.org
Open source digital forensics software.
URL: <http://www.sleuthkit.org/>

Snagit
A tool for capturing and editing screenshots.
URL: <http://www.techsmith.com/snagit.asp>

Wikipedia entry on Checksums
URL: <http://en.wikipedia.org/wiki/Checksum>