

✧ Introduction

Appraisal is defined in the *Collins English Dictionary* as ‘an assessment of the worth or quality of a person or thing.’ Archival theory extends this definition to include the policies and procedures used by an archivist to identify, evaluate and authenticate records, in all formats, which have transient or enduring value to record creators, institutions, researchers and society. Appraisal in a paper-based archive traditionally takes place once a record is no longer current, but determination of how long a record should be retained can take place before creation for some kinds of records. Examples of records often selected for permanent preservation prior to creation include the minutes of an organisation’s central executive or annual reports. Records chosen for indefinite preservation are usually selected on the basis of long-term historical significance. Records may also be selected for medium to long-term preservation for business purposes or to fulfil legal requirements; this chapter of the Workbook focuses on the approaches and techniques that can be applied in selecting records of long-term historical value. In a digital environment, traditional methods of appraisal need to be revised to take into account the need to archive objects before they become obsolete. It should be remembered that archival theories apply regardless of media.

Appraisal can never be free of bias and subjectivity, and will always reflect the culture and values of the day and the proclivities of the archivist(s) who made the decisions. Older manuscript material is scarce; this scarcity increases the value of surviving material and makes curators more inclined to keep a greater proportion of this material. By the late nineteenth and twentieth centuries, the age of ‘Modern Archives’, the documentary landscape had changed dramatically with the vast proliferation of both printed and archival material. Jenkinson, in his *A Manual of Archive Administration* (2nd ed. 1937), writes of ‘the increasing tendency to manufacture Archives on a hopelessly gigantic scale’ (p. 148). He attributes this to the comparatively cheap price of paper, the mechanical nature of writing and the tendency to ‘avoid the painful process of thought’ (p. 137) because using technology to make multiple copies of archives requires less intellectual effort than making a decision about the necessity of the copy. The scale of ‘Modern Archives’ required a regime of selection and destruction, which is not applicable to early manuscript materials. The American archival theorist, Schellenberg, was one of the first to address the need for wide ranging appraisal techniques and the destruction of superfluous records. Schellenberg worked at the National Archives and Records Administration for many years and his theories are based on public records, which were increasing at such a rate that a scientific method for selecting the small percentage with enduring value and destroying the remainder was required. Schellenberg’s division of appraisal attributes into primary (value to the creator - administrative, legal and fiscal) and secondary (historical value - evidential and informational) is as applicable to personal archives as it is to governmental records.

In the digital age the problems apparent in ‘Modern Archives’ are exacerbated. Digital abundance exceeds that of paper and tends to devalue individual items further. Creators of archives often keep more because creation, storage and discovery technologies are improving while appraising and disposing of material is still a largely manual and tedious activity. If these abundant digital assets are to be managed effectively the creator, the archivist and the researcher must do more in the way of appraisal.

Appraisal is the primary archival function on which all other functions depend and as such requires careful thought. Besides archival considerations, secondary factors such as the cost of storage, cataloguing and long-term preservation, will come into play.

Considerations affecting the appraisal of personal archives

Impact of personal record keeping behaviours

Different people have different ways of filing the same material. For example, at the offices of two Paradigm participants records were accessioned from two successive members of staff. In both cases, individuals who had the same job description as their predecessors and who created the same records series, used different file naming conventions. As in paper archives much depends on the record keeping behaviours of those creating the records. Logical arrangement and good file-naming conventions can assist appraisal, and more lax record keeping can significantly hinder the process. The arrangement of records within well-named series of folders allows the archivist to make appraisal decisions at the series level and to preserve existing meaningful relationships between items. Meaningful filenames make appraisal at item level simpler.

Appraisal timing

Traditionally, a politician's papers find their way to an archive towards the end of their working life, or they are deposited after their death by a relative. Quite often, papers are held untouched by the library or archive until staffing or funding considerations permit the creation of a full catalogue. This means that little or no appraisal is undertaken until immediately prior to cataloguing and, in most cases, cataloguing is delayed until a significant proportion of the collection can be opened to researchers. The same approach could be taken in a digital environment, but the need to create technical metadata for the preservation of digital objects at the point of ingest to the digital repository means that there is a strong case for bringing appraisal decisions forward, so that metadata is only created for objects that the repository will preserve. It might also be prudent to add high-level descriptive metadata (akin to box-listing) at this stage. At the very least, a collection level description will be required.

Financial and staffing constraints governing digital repositories may dictate that the creation of technical metadata and detailed appraisal cannot take place until funding is secured for a full catalogue description (usually when a collection can be opened). If this is the case it seems likely that *Archival Information Packages (AIPs)* will have to be created after only rudimentary appraisal and disposal. Towards the end of the digital archives record cycle, as preparations for opening a collection are under way, the archivist will have more time and knowledge to make a fuller appraisal of content. It is likely that the final detailed appraisal will take place during the cataloguing process when other records with no long term historical value will be identified and destroyed.

See the Paradigm diagram 'Digital Archives and the Records Cycle' (see p. 2) for an overview of the workflow stages and processes (including appraisal) undergone by digital objects from working with record creators, to acquisition and ingest into the digital storage area, through subsequent preservation strategies to the creation of a dissemination package for the end user.

Impact of collection development approach

The nature and timing of appraisal may also depend upon the collection development approach (see Chapter 02 *Collection development*) taken to acquire the papers. If snapshots of digital material are acquired during the creator's working life, there may be an opportunity to influence record keeping behaviours at the creation stage and to select only series of historical interest and possibly eliminate duplicate files prior to acquisition. Such an approach would minimise the extent of post-acquisition appraisal required and permit appraisal to take place soon after the materials are transferred to the Library. If the archive is acquired in a more traditional fashion, with the delivery of boxes of older papers, disks and other media, more appraisal time needs to be allocated. First, the archivist will need to judge whether it is worth the, sometimes considerable, expense of recovering archival material from media. Labeling may assist the archivist in this decision, but all too often labeling is inadequate to allow the archivist to form a realistic assessment of the media's contents. If the archivist proceeds to recover the data, it is also likely that some data will require migration to an accessible format so that the content and context of the records are accessible for making appraisal decisions.

✧ Introduction to disposal

The process of appraisal sometimes identifies material unworthy of permanent preservation; consequently, the issue of disposal and its documentation arises. Ideally, as with paper archives, the repository will have agreed with the depositor whether unwanted material should be returned to the depositor or destroyed. This may depend on whether the archive has been supplied with copies of material, which the creator also retains, or whether the archivist has acquired the depositor's last remaining copy.

Returning digital archives

If digital archives are returned, they should be returned safely and securely and a record of this should be associated with the collection. This return of archives might involve a reversal of the transfer process.

Destroying digital archives

If the depositor opts that the archivist should destroy the unwanted materials, then this should also be undertaken safely and securely and in compliance with the *Data Protection Act* where relevant. The complete destruction of digital data is not a simple act, but can be simplified if the repository manages the processing of the archives appropriately. Complete disposal requires repositories to understand where copies (deliberate, temporary, deleted and backed up copies) of files to be destroyed are located and how to completely erase the data. The ease with which inadvertent copies of digital data are created, leads us to suggest that digital archives be processed in stand-alone and secure environments rather than general day-to-day office computers, and especially not laptops, which are at greater risk of being stolen. Repositories should also limit and control the number of copies created very carefully.

Paradigm has not thoroughly investigated the topic of secure disposal, but would recommend that those undertaking destruction of sensitive digital materials do. Often the way in which media store data makes it very difficult to erase material completely. The success of computer forensics in recovering supposedly deleted data is evidence of this. Secure deletion mechanisms usually overwrite the data with randomly generated data to ensure that it is removed; several software utilities exist for secure deletion and Linux and Unix users may simply use the 'shred' command. As well as deleting copies from the repository's own system, the archivist may also need to destroy media supplied by the depositor. This might require special equipment or the services of a specialist contractor.

✧ Appraising digital records: a worthwhile exercise?

Is it necessary to appraise digital records at all?

Continual improvements to search and retrieval technologies and the apparently diminishing costs of storing vast quantities of material has led many to question the value of appraisal.

The arguments against appraising digital records:

Trends in the digital world seem to reject the practice of actively organising digital collections by choosing what to keep and what to discard. Declining storage costs and improved discovery seem to have rendered appraisal and disposal needless. Some record creators are increasingly reliant

on search and display technologies to present data: little is deleted and the creation of logical directory structures as a means of arranging material is abandoned. This approach to managing digital archives undermines the archivist's ability to appraise on the basis of content and context; effectively appraisal is impossible because the lack of structure leaves item-level appraisal as the sole option and the volume of the personal archive prohibits this. Appraisal could be postponed until the cataloguing process, when the archivist generates descriptive metadata (see Chapter 06 *Arranging and cataloguing digital and hybrid archives*) and identifies material requiring access restrictions for personal and legal reasons (see Chapter 09 *Legal issues*). It is also possible that the value of descriptive cataloguing for digital archives may be called into question. The needs of researchers may be satisfied by various search and discovery tools used in other digital scenarios; these may be preferred to, or used alongside, detailed and structured archival finding aids. Perhaps the only item-level descriptive metadata required by researchers (besides a collection level description setting out context, provenance and access restrictions, etc.) will be unique identifiers for citing specific digital manuscripts. It is possible that only the most important collections will be appraised and catalogued in future, similar to the way in which only very high value manuscripts are catalogued to piece level.

The arguments in favour of appraising digital records:

Arguments in favour of appraisal refute the argument that storage costs are declining. Storage and processing costs per byte may decline and continue to decline, but if the growth of content (per byte or per object) keeps pace with this declining cost, then the real cost of keeping everything may actually be the same as it is now, or higher. As storage and processing technologies improve, the technologies used to create data evolve to absorb this increased capacity. Further, the cost of backup routines and other system administration tasks have not decreased sufficiently that we can call storage 'cheap'. In addition to this, we must consider the costs of creating preservation metadata and undertaking preservation actions (e.g. format migration, which creates additional digital files and metadata for curation) for items of no historical value. Is the creation, storage and processing of preservation metadata sufficiently economic to offset the cost of not appraising? If automated tools for metadata creation and preservation actions are adequate, then perhaps it might be sensible to create technical metadata for all objects and combine appraisal with the cataloguing of an archive later on in its lifecycle. The timing and extent of appraisal in relation to a particular archive will probably be dependent on the arrangement of the material at the time of submission to the archiving repository.

✧ Issues to consider when making appraisal decisions

Personal record keeping differs from record keeping in an organisation or a government department in that the individual's day to day life does not revolve around a range of duties within a closely defined remit. Nor is the creator of a personal archive governed by records management strategies imposed across the board. This means that the individual's records are both more vulnerable to technological breakdown and much more idiosyncratic. Despite these differences, the characteristics to appraise in personal archives are similar to those used in an organisational context, though the results of the appraisal may be different.

Key characteristics to appraise

The first is the **content** of an archive. Does it contain records of long-term historical importance? Are the records what they purport to be? Can the records be manipulated and linked to others within the same digital or paper system?

Secondly, what is the **context** of the archive? Why were the records created? Do they have evidential value?

Thirdly, what is the **structure** of the archive? Does the structure shed light on the business, professional or organisational prerogatives of the creator?

Finally, there needs to be a **technical appraisal**. Can the collecting institution cope with the format of the digital records? Can they be extracted intact from their current hardware and software environment? Are the records supported by descriptive metadata? Are they valid instances of their format?

The relationship between cost and appraisal

Appraisal itself can be a costly exercise. Beyond initial appraisal, the cost and time involved in fully cataloguing a collection may also impinge on what is kept. The majority of detailed appraisal decisions are made at the full cataloguing stage, but appraisal and cataloguing at this level of the collection is expensive; it may be more pragmatic to retain everything and let the future researcher decide what is interesting. Free text searching within folders and across directory structures is possible, and techniques for searching other non-textual formats, such as audio and image may be developed. With such searching techniques it may not be worthwhile to catalogue below the collection and series levels.

Costs of preserving different types of digital object

Another key appraisal consideration is the cost of digital preservation. There is a need to weigh up the historical value of the record against the cost of indefinite preservation. Some digital objects may be more expensive to acquire, preserve and make accessible than others.

Objects that might be more expensive to retain include:

- Complex or compound objects, such as websites or email archives.
- Objects in undocumented formats.
- Objects in obscure formats.
- Objects in formats unsupported by a community or vendor at the time of acquisition.
- Objects in formats for which no migration/emulation tools exist.
- Objects in formats unknown/unsupported by preservation registries and tools.
- Objects for which the repository has no preservation strategy.
- Objects which are encrypted, password protected or subject to digital rights mechanisms.
- Objects on old or obsolete media.
- Objects without metadata.
- Objects which require software licences for access or manipulation.

Objects that might be less expensive to retain might be:

- Objects which are simple, such as a plain text file.
- Objects which are self-documenting.
- Objects in well-documented formats.
- Objects in formats which are supported by a community or vendor at time of acquisition.
- Objects for which good quality migration/emulation tools are available.
- Objects in formats which are known and supported by preservation registries and tools.
- Objects for which the repository has a preservation strategy.
- Objects which are not encrypted, protected by passwords or digital rights mechanisms.
- Objects arriving on contemporary media.
- Objects with good quality metadata.
- Objects which can be accessed and manipulated without the need for purchasing software licences.

Hybrid collections: managing both paper and digital archives

Many individuals are managing hybrid personal archives containing a mixture of digital and analogue materials. Even if a person creates archival materials digitally, often these are printed for use and annotated by hand. The offices of politicians participating in Paradigm created most records digitally, but printed some digital items to paper for use or filing. Archival materials accumulated by creators sometimes existed purely in paper format, for example, letters from constituents, invitations and press cuttings; equally certain materials were received only in digital format. Some offices had begun to digitise paper records and provide transcripts for the records using Optical Character Recognition software. Hybrid record keeping systems such as these risk unnecessary duplication which will need to be considered by the archivist if not the creator: a practical measure is to audit both the digital records and paper records together, establish where the same documents exist in both medium and decide which should be retained as the archival copy. Where both paper and digital copies exist, in the absence of manuscript annotations it would seem sensible to treat the digital as 'master' copy because the digital medium has functionality absent in the paper copy. Ultimately the decision must rest on the repository's ability to process and preserve the digital archive.

Factors which will need to be considered when appraising hybrid digital and paper records include how well the material is arranged, how much paper material there is and, more importantly, how easy it is to see how the paper material relates to the digital records in the same deposit. Understanding how the creator works can be very helpful in making these assessments.

✧ Appraisal-related issues encountered by Paradigm

Pre-acquisition appraisal

Paradigm started the process of appraising its exemplar personal digital archives with a records survey (see p. 24) prior to accession. The survey was designed to assess the content, context, structure and technical state of the creator's archives and was sent to participants in advance of the initial visit so that participants were aware of the kind of information required by the archivists.

During the initial survey, both the creator and archivist ruled out the accessioning of certain files and the archivist recorded information to assist in addressing various appraisal challenges.

Examples of files not selected in personal archives accessioned by Paradigm:

1. Casework files - some offices used specialist casework software for constituency work, which manages each case from creation until closure. This software allows the creation of reports on subject, ethnicity, age or gender of the constituent. Data can be extracted from the reports without revealing the identities of individuals. Paradigm's Academic Advisory Board argued that such survey data provided a snapshot of local political concerns of interest to social scientists, local and family historians of the future. Neither of the participating institutions have collected such material in the past, however, and given this lack of precedent, the volume of the series, and the huge Intellectual Property Right (IPR) (see p. 252) and sensitivity issues associated with it, Paradigm decided that casework files were beyond the scope of the project. Interestingly, the majority of Paradigm participants also expressed an unwillingness to permit access to casework files.

2. Public Records - Records created by ministers during the course of official business fall under the Public Records Act (see p. 247) and a small proportion will be selected for permanent preservation at The National Archives (TNA). Selection is carried out by the relevant Departmental Record Officer in conjunction with The National Archives' own selection criteria. The 'personal' aspects of politicians' private records are not classified as public records. However, where overlap of content occurs, it is necessary to synchronise cataloguing and access decisions with those taken for similar records held at TNA. Additionally, contact with the Cabinet Office may be required in some cases. Individuals prominent in other walks of life might find that records pertaining to their 'official' roles will be collected by local authority, organisational, and national archives.

Challenges associated with appraising digital archives

These are a few of the scenarios that Paradigm has faced in appraising its exemplar hybrid personal archives:

Managing large quantities of digital material

The key issue when appraising digital records, whether they be personal or organisational, is coping with the sheer quantity of information. It would be impossible to read and appraise the digital records of a twenty-first century politician document by document. Digital material can be copied, edited, and circulated rapidly. Such functionality inevitably encourages the proliferation of records.

Multiple copies of the same document stored by different people

Paradigm archivists often found that the same digital document was held by more than one of the politicians' staff members. In this scenario, the archivist must decide which version should be retained.

Large quantities of circulars

Mailboxes often contained a large number of circular-style emails, frequently with attachments, from the central office of a political party. In a paper archive such circulars are unlikely to have been kept but in a digital world there is an argument for keeping such working papers, which illustrate how the constituency MP fits into the broader political picture.

Multiple copies of the same document created during different snapshots

One of the key issues facing digital curators is how to deal with duplicate files especially if digital snapshots (see p. 11) are to be taken several times a year throughout a creator's working life. If such patterns of acquisition were adopted, there would be vast amounts of duplicate files: an overwhelming digital abundance. It will be necessary to use checksum or synchronisation tools to determine whether files with identical names are indeed identical.

System files present in acquisitions of digital records

The archivist will also want to identify and remove operating system, application and other software files (unless using the encapsulation approach to preservation, see p. 241). The National Software Reference Library¹ can assist with this; the NSRL provides a repository of known software, file profiles, and file signatures for use by those engaged in computer forensics.

Processing required prior to appraising digital records

Records accessioned on older media and/or in older formats will need to be extracted in order to undertake an assessment of their content, context, structure and technical viability. This can be challenging and may add significant costs to the processing of archives before appraisal even begins.

¹ National Institute of Standards and Technology, *National Software Reference Library (NSRL) Project website*. URL: <<http://www.nsrll.nist.gov/>>

Authenticity of author metadata in digital records

Dates

During both the appraisal and cataloguing process, decisions will be made based on the date on which a document was created, for example an urgent memo penned on the eve of a General Election is likely to have a greater significance than a circular written during parliamentary recess. Paradigm found that the concept of 'date' in a digital world is riddled with complications. The capture process itself can alter the perceived creation date. In practice the modified date is often the best indicator, as this shows when the file was last modified, while the creation date may only be that of saving the file to a new location. Further complications are introduced in office files where the creator has ticked the 'Update automatically' box when inserting a date and time field in the actual body of the text. This means that whenever the document is viewed the current date will be given. Paradigm encountered this issue in the press release documents generated by participants.

Authorship

Similarly, Paradigm found that the author metadata associated with many kinds of files is often inaccurate. This metadata can be particularly misleading in records acquired from under-resourced offices where staff routinely share computers. The poor quality of metadata also frustrates efforts to identify copyright holders in the digital archive; the process of circulating work can also lead to the decoupling of author and work. Establishing intellectual property rights will be a key concern for the digital curator who will need to determine who took a photograph or authored an article, whether they are still alive, whether they still hold copyright and how long that copyright will last.

Authenticity

Authenticity is another key consideration. When appraising records the digital curator must establish that they are indeed what they purport to be. If authorship of records is in doubt, the value of the records diminish and they may be unsuitable for long-term preservation as archives. Most creators of office documents do not take pains to make sure that they are automatically described as the author by their software package and do not add their name in the text of a document or incorporate a digital signature. Authors often rely on the means of circulating the work, such as an email, to assert their authorship, but if the means of circulation and the document are disconnected, the provenance is undermined. Microsoft Word does keep a 'revision log' which provides some additional metadata, but accessing this is not as straightforward as accessing the file properties dialogue. The native metadata associated with emails is much more reliable and accessible as the email header provides a great deal of information, which is accurate and complete in most cases. Establishing the authenticity of archives repositories wish to purchase will become increasingly important and may be complex if material includes examples of older or obscure technology.

Relevance of format to appraising digital records

The format of a document may assist the digital curator in assessing the purpose of a record and thus its value. For example, a document saved in PDF implies that it is the final version of a report, which has been passed as fit for public consumption. By choosing to save in PDF the author of the document is consciously preventing further copying or editing of a document and in this respect maintaining its authenticity. Similarly, a snapshot of a politician's personal website acquires a final and authoritative version of the site as it stood at the moment of capture. Arguably, researchers may be more interested in the less sanitised picture - the content rejected for the public arena or the drafts that led to the final versions. Formats used for the drafting process, such as word-processing formats, may therefore embody high value archival materials.

Paradigm's Academic Advisory Board identified email as one of the most interesting types of historical record being created in our times. It contains records of business transactions (that might have been undertaken via an exchange of correspondence on headed notepaper previously) as well as informal exchanges (previously the stuff of telephone conversations). The variety of functions served by the email format would suggest that making appraisal decisions on the basis of format alone is insufficient.

Comparisons with appraising paper archives

Arguably, many of the issues described above are also present in the paper archive.

Managing large quantities of material in paper archives

Paper archives can be vast and include several copies of the same document with slightly different provenance, or of different versions of a document. Snapshots of paper personal archives are also acquired, although accessions are typically less frequent and regular so the likelihood of introducing duplicates is perhaps smaller.

Processing required prior to appraising paper archives

The need to process some digital materials prior to appraisal is also mirrored in the paper world where the intervention of a conservation expert may be required prior to full appraisal.

Establishing the authenticity of creation dates and authorship in paper archives

Questions regarding authenticity and context apply equally to paper archives: the dating of documents and identification of creator has always caused problems. Handwriting is an obvious clue missing from the digital record yet paper archives present similar problems of authorship, for example, typed memos where the author is not given. In archives and libraries across the world there are numerous photographic prints without known date, subject or creator. Researchers are used to dealing with such uncertainties in the paper world, yet the expectations for digital records are higher. Researchers expect accurate creation dates as, on the face of it, they appear to be generated precisely and automatically by the computer. Archivists and researchers have established tools and techniques for scrutinising paper records, but may need to learn additional skills to work with their digital equivalents.

Relevance of format to appraising paper records

Like digital records, the purpose and audience of some paper records can be discerned from their format (e.g. a note written on paper from a spiral notepad will be more informal than a memo on headed paper) and may provide information useful to the appraisal process.

✧ Recommended approaches to appraising personal archives

Functional appraisal theory

Paradigm's experiences suggest that the digital curator would be on safer ground assessing the function not format of a document as part of the appraisal process. Two archival theories which approach appraisal in this way are macro appraisal and functional appraisal.

Macro appraisal

Macro appraisal is a 'top-level', systematic approach to selecting which records have long-term historical significance. It is often used by national archives to appraise the great swathes of governmental records from which a mere three percent will be selected for permanent preservation. The quote given below gives a useful definition of what is meant by macro appraisal.

“Essentially, macro-appraisal shifts the primary focus of appraisal from the record - including any research characteristics or values it may contain - to the functional context in which the record is created. The main appraisal questions for the archivist are no longer what has been recorded, where it is, and what research value it has. Instead, the archivist uses knowledge gained by a functional analysis of an institution, including an analysis of the interaction of function and structure, of organizational culture, of records keeping systems, and of citizen-client involvement with the institution or function.”

Source of quote: *Appraisal Methodology*. Library and Archives Canada, Information Management Services, Summer 2000.

In the context of political papers macro appraisal would start with the broad area of UK politics, identifying the most important organisations (e.g. House of Commons, House of Lords, political parties, and think tanks) followed by the key functions (e.g. policy making, law-making, constituency work, and treasury), the record creators in each function (e.g. politicians, politicians’ secretaries, think tank employees, and members of committees), and the key political events (e.g. elections, Prime Minister’s Question Time, debates, and votes). Individual politicians would have a place in this structural map probably at the same level as a party political organisation. The very act of mapping politics in this fashion instructs the archivist in the political process and will alert them to key political documents such as Hansard (verbatim record of speeches made in Parliament), case work files, websites for elections, etc. Armed with this knowledge, the archivist will have a good idea of which are the most valuable records and be able to target their collecting accordingly.

Functional appraisal

Functional appraisal is a subset of macro appraisal, which can be a useful approach for digital records. Again, instead of looking at the document itself the digital archivist looks at the functions carried out by the record creator. This allows the archivist to select records for preservation based on the functions that generated the records not the content of the record. Functions are more stable over time than naming and filing practices and less susceptible to technological changes.

Macro and functional appraisal were developed with reference to modern governmental agencies and have been applied to organisations not individuals. Yet arguably such approaches lead to a greater knowledge of the subject area and a propensity to seek certain records rather than purely accepting what is offered, thus maximising the quality of material when resources are of a premium.

Both macro analysis and functional analysis played a role in the selection of participants for the Paradigm project and in the initial records survey (see p. 24) used by the Paradigm team. Many of the questions asked in the records survey were about analysing functions and identifying the record-types used in carrying out these functions. Perhaps the most sensible course for a digital archivist dealing with hybrid or digital personal archives would be to use macro appraisal or a ‘top down’ approach during initial selection and processing (incidentally this would be the stage to eliminate duplicates too), but to use a ‘bottom-up’ strategy when carrying out more detailed appraisal during the preparation for cataloguing.

Bottom-up appraisal

A more granular assessment of an archive’s contents will typically be undertaken when surveying a collection immediately prior to cataloguing. It is rarely cost effective to weed out single documents from a folder whether in a digital or paper environment. Not only is this a waste of time it also disrupts the context of a record. By taking a sample of folders, it should be possible to assess the reliability of folder and document names used by a particular creator. Where records series have been satisfactorily assigned meaningful titles it will mean that appraisal can, for the most part, operate without having to open each folder. A document-by-document approach to appraisal should be taken only as a last resort or when the digital archivist is faced with an example of terrible record keeping.

Technical appraisal

A digital repository will need to develop ingest workflows for common file types which implement the preservation strategy (see Chapter 08 *Digital preservation strategies*) adopted for those file types. The appraisal process will therefore include a survey of the types of digital records that have been accessioned. Depending on the strategy adopted, it may be necessary to migrate less supported file formats to those that are more likely to endure over time. There are various tools and registries which can assist in determining file types and which provide information upon whether a specific file format is supported or likely to become obsolete. In some cases data extraction work will be required before a complete appraisal of content, context and structure will be possible.

✧ Practical solutions advocated by the Paradigm Project

Encourage best practice at the start of the record cycle

Archivists working with businesses, organisations and governmental bodies are encouraged to become involved in the design and implementation of electronic record keeping systems. This is not an option for institutions dealing with the records of individuals. However, it is possible that projects such as Paradigm can raise awareness of the issues through the publication of guidelines (see *Appendix B: Guidelines for creators of personal archives*) and by encouraging the appropriate professional bodies to issue codes of practice. If creators can be encouraged to curate their own personal digital archives to some extent (including some basic appraisal), then the appraisal of personal digital archives by archivists at the time of deposit will be much simpler.

Sample advice archivists might offer creators

Digital photographs

Ask creators to save photographs in folders rather than randomly in 'My Pictures' folders. If all pictures from a visit to Berlin in April 2005 were stored in a folder called 'Berlin 2005' this would assist the archivist immensely, even if the photographs retained the filename generated by the camera (usually a random, meaningless string of characters). If the creator wanted to be more helpful, they could include a plain text file in the folder, which included details of the visit to Berlin, when it occurred, why it occurred and whom it involved.

Emails and Office files

Creators should be encouraged to file their emails and documents in folders with meaningful titles and to delete routine emails and documents that obviously have little value.

Obtain contextual information from record creators

The records survey assists the archivist in understanding the functions undertaken by the politician and their staff. It also generates information about the kind of records created by the office and the record keeping practices employed. In addition to surveying and interviewing key records-creators, the survey process also includes obtaining documentation created by the offices themselves, such as manuals for archiving systems, or job descriptions. This background information on roles, tasks and procedures will be useful during the appraisal and the cataloguing processes. Such information is also likely to be of interest to future researchers.

Eliminate duplicates

Duplicate documents should be eliminated by comparing the checksums of digital objects.

Eliminate standard operating system and application files

Generally speaking, application and standard operating files (e.g. Clipart) can be routinely destroyed as such digital objects are not archival or pertinent to the personal archives of an individual

and may require licences to retain. Archivists may wish to retain some software (where licences permit) to assist in future data extraction processes.

n.b. The ‘encapsulation’ approach to preservation (see p. 241) would recommend the retention of all system files.

Practical tests

Photographs

The popularity of the digital camera has vastly increased the number of photographs found in the average computer directory and on mobile devices and web services. The ease with which images can be taken and the fact that images can be captured and discarded at minimal cost has encouraged their exponential growth. Archivists familiar with cataloguing paper photographs will know how difficult it is to identify individuals correctly and to deduce the date, place and occasion of a photograph unless such descriptive data has been recorded on the reverse of the image. The same is true of digital images unless meaningful metadata has been recorded at the time of capture or storage. Most individuals do not record in detail the ‘who, why, when and where’ when taking a digital photograph as the image has been taken for their own use and they are likely to know this information. Digital cameras automatically generate a limited amount of metadata such as: dimensions, date, camera model, type/format, size, but the standards used to record this metadata (and the metadata contained in those standards) vary from manufacturer to manufacturer and will need to be converted to a standard descriptive and technical metadata schema. Such technical information will be vital for the technical aspects of digital preservation but has limited use for descriptive cataloguing purposes.

Appraising photographs found in Paradigm accessions

One sample accession included a series of images recording a dilapidated building in Manchester. The folder was named, rather unhelpfully, ‘Manchester building’ and held six untitled images, aside from the automatically generated date there is little else to go on. This was a classic example of the problems associated with appraising digital photographic collections and later creating meaningful descriptive metadata for images.

Paradigm found that file directories often contained many images taken of the same composition to maximise the chance of a good shot. Where numerous images of essentially the same shot exist, there are grounds for selecting the most representative and disposing of the rest.

Another more useful example from one collection is a folder holding an image and a related press release. The transfer process obliterated the date the image was taken but it is likely to have been taken around the date of the press release, 11 July 2003. The image showed an alarmed politician wearing a target and undergoing a mock execution by a masked man wearing khaki. Without the accompanying press release, a cataloguer may have been at a loss to explain what the image documented. From the press release it becomes obvious that it was a publicity stunt in which the politician along with 12 trade union general secretaries, and a former minister were ‘mock’ assassinated outside Parliament for protesting against the UK Government’s continued military assistance to Colombia - not something your average archivist could have guessed!

Many images will have originally arrived attached to an email. The ‘carrier’ email is likely to contain some descriptive metadata relating to the image. It may be possible to find the date when the image file was transferred and search for all incoming email for that date. Such a search might generate additional contextual information. Conversely it is likely that some images will only exist as attachments and will not have been saved and stored elsewhere in the computer directory. Again much will depend on the record keeping practices of the individual creator.

One clear advantage of working with contemporary politicians, and dealing with material likely to be less than 5 or 6 years old, is that it is possible to ask the creators of the photographs to provide missing contextual information. For some creators, this might mean working through the photographs image by image; others may only have the time to assign a date or an occasion at a group level.

Appraising an email directory

Paradigm found that the way in which an archivist could approach the appraisal of an email directory was very much dependent of the way in which the directory had been structured by its creator. Some depositors arranged emails into logical folders and others did not. For structured directories a folder-by-folder appraisal can be undertaken; this approach takes the title of folders into consideration and perhaps examines some sample emails within. The appraisal of email directories with very little structure is a far more daunting task. It is impossible to read each email and judge its worth on an email-by-email basis, though sorting and searching techniques might allow the archivist to delete, or be sure to keep emails from certain individuals and organisations, or concerning a certain subject. In such circumstances it may be best to preserve the entire email directory, subject to depositor restrictions, without undertaking any appraisal. In this scenario, the future researcher would be reliant on the search and sort facilities of access mechanisms to make productive use of the email archive.

Another question that may arise when appraising email concerns the retention of spam email in an individual's personal archive. There may be a case, for some institutions at least, to collect electronic ephemera, such as spam, as historians of the future might find it to be of great interest in the same way as scholars have found a wealth of information on quack doctors and pseudo-scientific remedies from the small advertisements in the popular press of Victorian times.

Paradigm found that there were some ethical questions to address in the appraisal of email too. These included the retention of deleted email found in depositors inboxes: creators did not necessarily realise that an email archive would include these emails. For email archives where select folders are to be given, there is also the question of sentmail, which is typically not arranged to mirror any directory structure in place for incoming mail and may include email relevant to incoming email not accessioned as part of the email archive. Ideally the archivist should ask the depositor whether such emails should be retained as part of the archive.

Developing tailor-made appraisal criteria

Aside from straightforward rules such as removing duplicates, and material that has no long-term research potential, each professional group will have its own concerns. The types of records created by politicians will be quite different from those created by an author or scientist. Selection criteria should be subject to regular review - record series may change over time and new types of documents may become significant. Examples of evolving records include personal websites/blogs and personal digital video, often stored by web services.

Paradigm found it useful to consider the paper records of politicians held in Oxford and in Manchester in order to scope the structure of hybrid personal archives (see p. 19); this information was used alongside the records survey of digital archives to establish key records series for politicians.

Main record groups found in a politician's office:

- Biographical information.
- Central party circulars.
- Correspondence.
- Engagement diaries.
- Election material.
- Financial records.
- Manifestos.
- Photographs (digital and hard copy).
- Policy documents.
- Press releases.
- Press cuttings.
- Printed matter.
- Reports.
- Speeches.
- Staffing (job descriptions are useful the rest can probably be discarded).
- Subject files.

Files which should not be accessioned:

- Casework files.*
- Cheque book stubs (paper).
- Copies of personal expense claims.
- Ephemera (paper).
- Membership databases.*
- Miscellaneous newspaper cuttings (paper).
- Office templates.
- Personal files of office staff.
- Routine requests for information.

*There are good arguments for keeping such information, which would be of interest to local and family historians, but it was deemed beyond the scope of Paradigm.

✧ Useful appraisal tools

Paradigm experimented with a number of tools and approaches to assist in the appraisal process. These included tools for generating directory structures, directory management tools, synchronisation tools and digital forensics tools.

Directory tools

Paradigm experimented with tools such as DirPrinting¹ and Karen's Directory Printer,² which can be used to generate a complete list of all folders, subfolders and files in a directory providing the archivist with an overview of the accession. DirPrinting offers a drive/directory window and a file list, which can be arranged and sorted according to various criteria. This list can be sent to any printer, copied to the clipboard and pasted into an application, saved as file, and sent as an email. Karen's Directory Printer, as the name suggests, gives the user the option of printing out the directory or saving to a disk as a simple text file. The software has options for sorting the directory by file name, extension, file size, date created, date last modified, date last accessed. The usefulness of these tools can be limited if files and folders have been assigned poor titles, but the process should identify where data and system files reside.

1 DirPrinting. URL: <http://www.majusoft.de/DirPrinting/index_en.htm>

2 Karen's Directory Printer. URL: <<http://www.karenware.com/powertools/ptdirprn.asp>>

Directory management tools

Directory management tools such as Windows Explorer Browser, a standard feature of Microsoft Windows, can be helpful during the appraisal process. By selecting the 'details' option from the 'views' menu it is possible to see details of each document, including title, size, document type, date created, etc., but not to output these in a useful way. Users can also select which bits of metadata to view (if those metadata exist for a particular file) from a dialogue box located at 'View' > 'Choose Details..'. An example would be to view all the images in the 'My Pictures' folder. For each image, useful descriptive and technical metadata is displayed, though it should be remembered that the metadata does not necessarily provide accurate information about creation date and creator.

Synchronisation

Paradigm experimented with synchronisation tools as a means of identifying duplicates within and between accessions. Tools such as Unison¹ (a file synchronisation tool for Unix and Windows) are able to compare two sets of files and identify whether the individual files have changed or if they are identical to the other set. This is a useful tool for appraising a series of snapshot accessions to determine which, if any files, have changed.

File comparison tools

Other useful tools include 'compare features' functions, such as that of OpenOffice, which allow the archivist to examine two documents (perhaps two with the same name) to determine exactly where they differ.

Digital forensic tools

Paradigm utilised some digital forensic tools as a means of exploring digital archive accessions. These provide similar features to directory tools, but often include additional functionality, such as the ability to preview certain types of file in the forensic application, or to export a list of file metadata, including file path and checksum, in CSV format. Paradigm found this metadata export very helpful, and used it to generate a spreadsheet to manage the appraisal and arrangement of the digital components of the personal archive of former cabinet minister Barbara Castle.²

Useful resources

Cunningham, Adrian, 'Waiting For The Ghost Train: Strategies For Managing Electronic Personal Records Before It Is Too Late', *Archival Issues: Journal of the Midwest Archives Conference*, 24, 1 (1999), pp. 55-64.

URL: <<http://www.mybestdocs.com/cunningham-a-pers-er.htm>>

Dalgleish, Paul, 'The appraisal of personal records of members of parliament in theory and practice', *Archives and Manuscripts* (May 1996), pp 86-101.

Committee on Appraisal, International Council on Archives, *[Draft] Guidelines on Appraisal* (2003).

URL: <<http://www.ica.org/en/node/30417>>

DirPrinting

URL: <http://www.majusoft.de/DirPrinting/index_en.htm>

Forensics Wiki

URL: <http://www.forensicswiki.org/wiki/Main_Page>

1 Unison File Synchronizer. URL: <<http://www.cis.upenn.edu/~bcpierce/unison/>>

2 Forensics Wiki, *Forensics Wiki website*. URL: <<http://www.forensicswiki.org/>>

Jenkinson, Hilary, *A Manual of Archive Administration*, 2nd edition (London: Percy Lund, Humphries & Co., 1937).

Karen's Directory Printer

URL: <<http://www.karenware.com/powertools/ptdirprn.asp>>

Library and Archives Canada, Information Management Services, *Appraisal Methodology: Macro-Appraisal and Functional Analysis* (Summer 2000).

URL: <<http://www.collectionscanada.ca/information-management/007/007007-1035-e.html>>

O'Shea, Greg, 'Keeping Electronic Records: Issues and Strategies', *Provenance the Electronic Magazine*, 1, 2 (March 1996).

URL: <<http://www.provenance.ca/1995-2000backissues/vol1/no2/features/erecs1a.htm>>

Paquet, Lucie, 'Appraisal, Acquisition and Control of Personal Electronic Records: From Myth to Reality', *Archives and Manuscripts* (November 2000), pp 71-91.

Piggott, Michael, *Appraisal - the state of the art* Paper delivered at a professional development workshop presented by ASA South Australia Branch (26 March 2001).

URL: <<http://www.archivists.org.au/appraisal-state-art-26-march-2001>>

Schellenberg, T.R., *Modern Archives: Principles and Techniques* (Melbourne: F. W. Cheshire, 1956).

Schellenberg, T.R., *The Management of Archives* (Columbia University Press, 1965).

Thiele, Harold E., 'Appraisal, Provenance, and the Computer Revolution: An Examination of Organizational Records in the Electronic Age', *University of Pittsburgh Katharine Sharp Review*, 6 (Winter 1998).

URL: <<http://mirrored.ukoln.ac.uk/lis-journals/review/review/6/thiele.pdf>>

Unison

URL: <<http://www.cis.upenn.edu/~bcpierce/unison/>>