

✧ Introduction

Collection development is an umbrella term for the approaches taken by archives, libraries and museums in developing collections for their end users. It is usual to publish a collection policy which outlines the areas in which the institution collects and which expresses collection development priorities. Institutions also employ collection development strategies to fulfil the aims of a collection development policy. Such strategies are not necessarily public documents, or even documented at all, but they represent the approaches by which an institution will acquire new materials to enhance its collections. Collection development strategies should consider the issues around the collecting of archives and develop approaches which reduce the risk of failing to meet the aims of relevant collection development policies and provide the best opportunity of acquiring appropriate and high value collections for users of the repository.

Traditional collection development strategies have included:

- Assessing archives that are offered as gifts, on loan or for purchase.
- Actively monitoring auctions and sales.
- Actively pursuing archival collections directly with creators who are well-established as historically significant.

The risk to digital archival materials is that degradation or obsolescence could render them inaccessible, or a complex processing prospect, long before they reach an archive via these routes. This could become increasingly true as those born today may use evolving technologies throughout their lives, producing personal archives that represent several generations of computing technology and practice.

Drawing on its experiences, Paradigm proposed a series of possible approaches to developing collections of born-digital (or hybrid) archives in the belief that pursuing traditional approaches alone risks that institutions collecting personal archives will fail to fulfil the ambitions of their collection development policies and the quality of collections documenting the contemporary period for future research would diminish.

This chapter of the Workbook explores Paradigm's proposed approaches to collection development. These aim to address issues prominent in the collection of personal digital archives, such as media degradation, various kinds of technical obsolescence, distributed storage, abundance and poor structure. Six basic approaches to the development of personal archives including digital materials are described:

- Regular snapshot accessions.
- Post-custodial approach.
- A combination of the snapshot and post-custodial approach.
- The traditional approach.
- Periodic transfer of records via retired hardware and media.
- Self-archiving.

✧ Regular snapshot accessions

In this approach, the archivist makes regular snapshot captures of digital records which are exact copies of the data held by the creator and which are immediately accessioned into a managed digital repository. Complementary paper records could also be accessioned, where appropriate, during the same visit. The visiting archivist would also advise the creator on the creation, management and storage of both their paper and digital records. It is important to reach an agreement with the depositor about the kinds of materials that can be accessioned in these snapshots, and to decide on a frequency and timing of snapshots that is appropriate to the quantity of material being created and the working year of the creator. The snapshot approach to collection development is often applied when collecting the archives of organisations. This periodic sampling gives a snapshot of the organisation at regular intervals and it is interesting to see the rise and fall of relationships between individuals and organisations from one accession to the next. This would be equally interesting in the context of personal archives.

Advantages

- Regular meetings help to build a positive relationship with the creator leading to a high level of trust and a greater willingness to transfer more sensitive records. Regular contact leads to greater commitment on both sides and a stronger likelihood of ultimately acquiring the complete archive.
- Accessioning records earlier in the lifecycle will increase the chances of longevity for some objects.
- Accessioning near contemporary records can make the appraisal process much easier because the events and personalities (major and minor) to which the records pertain will be fresher in the archivist's mind. The archivist would also have the advantage of being able to consult the creator for contextual information.
- Regular visits allow the archivist to keep in touch with the record keeping practices of the record creator and offer timely advice.
- Many visits over the lifetime of the record-creator will provide the archivist with excellent contextual information which will inform the descriptive metadata and in turn aid the eventual discovery and usage of records by researchers.
- Greater likelihood of capturing early drafts/versions of documents, which might be refined later on and earlier versions overwritten.

Disadvantages

- Collections composed of frequent accessions may become over-large. Improved search and discovery tools and declining storage costs may undermine this argument for digital objects, but not for the paper materials which accompany them.
- Appraising collections sooner after creation may be more difficult. Often events and personalities only become major (or minor) in retrospect, and there may be a danger of destroying something which could have huge significance later on.
- Because copies of digital material are taken, it is possible that some files may be accessioned more than once, resulting in the duplication of some files across snapshots. This will need to be considered when developing appraisal and description guidelines and strategies may need to be developed to deal with issues such as versioning and duplication.
- Changes in the personal or working life of the individual may make such a relationship difficult to sustain.
- The increase of contract staff in archives may present difficulties in sustaining long-term relationships of this kind.
- The creator might view the digital repository as a records management service and devolve the management of current and semi-current records to the Library, including storage and retrieval.
- Regular visits from the archive can be perceived as inconvenient. Accessioning large amounts of digital material does take time.

- If a third party is administering the process on behalf of the creator responsible for the records (e.g. the assistant of a politician), it can be difficult for them to obtain the necessary approval from the creator to transfer the records.
- The creator may wish to transfer contemporary analogue materials alongside digital materials. This could result in a substantial paper archive, which must be stored for some period of time before it may be accessed by researchers.

✧ Post-custodial approach

The archivist develops a continuing relationship with the record-creator and provides advice over a number of years. Such advice is likely to include the creation of regularly updated basic guidelines on best practice management of digital records as well as advice given in response to personal preferences relating to creation or storage. The aim of this approach is to allow the creator to retain custody of the archives, but to assist them in managing their records so that they can be transferred in good condition when the timing is right.

Advantages

- A creator with confidential records may be willing to deposit records at a later date when they are less sensitive and when their public exposure would be less damaging. This approach allows the archivist to provide guidance that increases the chances of these records surviving in accessible form so that they can be accessioned at a later date.
- The approach should increase the chances of bit level preservation. This is where the digital object survives intact, but may not be easily accessible due to software or hardware obsolescence.
- This approach ensures that the creator has access to, and remains responsible for, their own records and delays the burden of preservation for the digital repository.
- The records may have undergone some appraisal by the records creator before transfer to the archive. This will reduce the bulk of the collection.

Disadvantages

- In the absence of scheduled accessions, it is possible that the creator will see no reason to keep in regular contact with the archivist, so the relationship could lapse. This would be particularly true when the repository is reliant on intermediaries, such as politician's assistants, which may have high turnover, or where staff turnover is high at the repository itself.
- The difficulties of sustaining the relationship through change in the creator's life and change at the repository described in the 'snapshot' approach above also apply to this approach.
- Without any form of legal agreement, the creator is under no obligation to place their archive with the repository. Sustaining a relationship of this kind is an investment, and if the collection were given elsewhere, destroyed or sold, this investment would be lost. An agreement could be drafted, which would allow the Library to recoup its investment in such scenarios, but would need to be worded in such a way that the creator would not be penalised in the event of unforeseen catastrophic data loss. The creation of an agreement like this would be difficult because the Library is dependent on the goodwill of the creator.
- As the archivist cannot ingest the materials identified for long-term preservation into a managed environment it is difficult to create granular preservation metadata. Such preservation metadata could be used to monitor the status of the collection and to advise creators when objects in their collection are 'at risk' of obsolescence and to conduct preservation actions in a timely manner.
- Although well intentioned, it is probable that most creators of personal digital collections have neither the time, expertise or technical support to care for digital records selected for long-term preservation.

✧ Combining snapshot & post-custodial approaches

This approach combines the ‘snapshot’ and post-custodial approaches and offers a mix of the advantages and disadvantages applicable to these approaches. The approach allows archivists to make regular accessions of selected digital materials so that they can be ingested into a managed environment. It also allows the creator to retain sensitive series of digital records that they would prefer to transfer later, but to benefit from the advice of the archivist in caring for that material. Unlike its digital equivalent, the hardcopy element of a personal archive will not require regular preservation attention. It therefore makes sense to defer the accession of paper-based materials; this may also give creators time to undertake some basic appraisal of the material. Combining snapshot accessions of digital material with post-custodial collection of sensitive digital and all hardcopy components is probably the most realistic option for many circumstances. It allows at least a portion of the digital collection to be managed in an OAIS-compliant digital archive, while offering a better chance of survival for sensitive contemporary digital records that creators are not yet ready to transfer.

Advantages

See snapshot (p. 11) and post-custodial (p. 12) approaches.

Disadvantages

See snapshot and post-custodial approaches. It is hoped that a significant amount of digital material could be accessioned in this joint approach which would lessen the weaknesses apparent in the post-custodial approach.

✧ The traditional approach

Repositories approach creators, or creators approach repositories, at the end of their careers/lives when they are more engaged with their history and the sensitivity of their records have decreased through the passage of time. It is probable that this approach to collection development will always exist alongside other approaches.

Advantages

- Creators are likely to be more enthusiastic about the transfer of their records to an archive.
- Creators may have more time to participate in the process.
- It is likely that the creators represented in the archive will be older individuals, therefore protections enjoyed under the *Data Protection Act* and copyright legislation will expire sooner than with material accessioned earlier after creation.
- It is likely that these collections can be made publicly accessible much sooner than archives accessioned earlier in the lifecycle.

Disadvantages

- This approach is likely to entail the loss of many digital records. Even records which appear to have survived at the bit-level may not be accessible.
- The approach may require some ‘digital archaeology’, which can be unpredictable and expensive.
- Where records have undergone migration from one computer to another it may not be possible to show that they are authentic and have not lost some of their original attributes.
- There is no opportunity to learn about the context of the archive in its contemporary setting.

✧ Transfer via retired media

In this approach, the archivist maintains a relationship with the creator, as in the 'post-custodial' and 'snapshot' approaches, but only accessions records when the creator is finished with them or the media or hardware on which they reside. This means that the archive should receive a series of retired computers and old media during the life of the creator. In the case of hard disk drives, the accession will include operating system(s), applications and software libraries, etc., as well as the records of the creator. The archivist will need to decide whether the media and hardware deposited ought to be retained, and for how long. If the media and hardware are not to be retained, the archivist will also need to consider whether software files relating to the deposits ought to be retained.

Advantages

- Records are transferred in their original containers, which ensures that their original order survives intact.
- Records are transferred sooner than in the 'traditional approach', increasing the likelihood of bit level preservation and preservation of access.
- This approach should reduce duplicates as the archivist is accepting the record on the creator's own storage devices rather than taking copies from these devices, as in the 'snapshot' approach. It does not guarantee that the copy of the record is the sole copy, however.
- Transfer of records on hard disk drives increases the likelihood that the archivist possesses at least one device with software that can read the creator's archives in their original formats.
- The archivist should have regular contact with the creator, which allows advice to be given and for the archivist to learn about the context of the records.
- The records may have undergone some appraisal by the records creator, reducing the bulk of the collection.

Disadvantages

- The archivist must make decisions about the software, media and hardware associated with an archive. Ought it to be preserved or disposed of?
- It is possible that the creator may turn to the archivist for salvage of older digital records. Sometimes it is not clear whether data on the storage device is of historical value, or whether it is recoverable at all, prior to commissioning expensive recovery.
- When computers are retired, it is often because they are failing or have failed. This means that recovering the data could be precarious and loss may have already occurred. This approach requires the archivist to have access to hardware, software and expertise which can recover authentic records from older media, or an adequate budget (and the creator's agreement) to outsource this activity. Even if the latter approach was adopted the archivist would require a considerable amount of technical expertise to be able to accurately identify the issues and put the project out to tender.
- For the archivist to use, copy or retain any proprietary software on the hardware it is likely that the owner of the end user licence must transfer this licence to the archive. Not all software licences give users the option of transferring their licence to another person or organisation, and some that do include restrictions. Many licences will also require that the licensee delete software from old hardware after moving it to new hardware and do not specify an end-date when some/all of the licence's restrictions cease.
- The archive could be overwhelmed by older physical media which have new storage requirements and will occupy a great deal of physical space.
- The approach relies partly on the creator's right to transfer hardware and software to the library. In some cases personal records will be stored on equipment owned by the employer.
- There are potential health and safety issues associated with the retention of degrading hardware which include heavy metals and other potential contaminants.

✧ Self-archiving

Using this approach, heritage institutions would provide digital services to individuals whose personal archives interest them. These services would help individuals to create and use their content, while allowing the Library to bring this information into a managed environment. Relevant services would include email and chat, blog tools and hosting, weboffice facilities, personal digital assistants, data storage and backup. Alternatively, the Library could offer a more basic repository service, which allows users to upload their personal digital collections remotely. In this scenario, the creator could be responsible for creating their own initial descriptive metadata and may need either training or guidelines to encourage consistency of approach.

Advantages

- The provision of tools and services mean that creators can be encouraged to use open, preservation-friendly, formats.
- By managing the content for the creator's contemporary needs, the archivist can also monitor the status of digital objects and undertake preservation actions as and when required.
- The creators may be more likely to embrace such a system as it not only preserves material for posterity but in the near term gives creators the security of having a backup copy saved elsewhere.

Disadvantages

- Users choose software and services for various reasons: reliability, functionality, branding, defacto - it came pre-installed, familiarity, etc. Heritage institutions have little experience in this area and would be competing with organisations, such as Google, whose primary business is the provision of software and services.
- A service aiming to supply tools and services for users beyond a secure upload would require substantial investment to establish and maintain.
- The service would need clearly specified service level agreements and it could be difficult to offer the flexibility needed when dealing with personal collections.
- There are legal considerations that need more thought. It may not be possible for individuals to use the service for some of their personal records. For example, politicians could not upload records governed by the Public Records Act.
- The service may conflict with services provided by an employer.
- Other creators contributing to the individual's personal archive could object.
- The creator might view the digital repository as a records management service and devolve the management of current and semi-current records to the Library, including storage and retrieval.
- It is likely that, at least initially, creators would not place extremely personal or sensitive material in a 'self-archiving' repository. In time, as the trustworthiness of such a scheme was proven, ingest of sensitive material may increase.

✧ Conclusions

From the analysis of the six basic approaches presented above it is evident that there is no simple solution to successfully developing collections of personal digital archives. To date, methods of collecting personal digital archives are largely untested. A pragmatic approach which combines 'snapshot', 'self-archiving' and 'post-custodial' collecting with the transfer of retired hardware and media in the case of very eminent individuals may offer the best solution at this point in time. What is apparent is that the approach will need to fit with the preferences, personality and skills of the individual creating the archive. It is also clear that the earlier personal digital records enter the managed environment of a digital archive the greater their chance of permanent survival. How-

ever, this has to be measured against the information overload (and likely abundance of low value material) threatened by the snapshot approach versus the more selective (and hopefully higher quality) material generated by 'self-archiving' and 'post-custodial approaches'. The latter reach the managed environment of a digital archive at a later stage but are likely to have undergone partial appraisal by the creator.

General issues

Paradigm also discovered some more general considerations potentially relevant to several collection development approaches. These include:

Legal agreements - defining expectations

At the outset, regardless of which approach or approaches are adopted, collecting institutions should draw up a legally binding contract which clearly establishes what is expected from both the creator and the collecting institution. Such an agreement will protect the collecting institution from investing time, energy and resources in a collection which is ultimately deposited elsewhere. The deposit agreement will also need to address data protection, privacy and copyright restrictions.

Defining 'advice'

It is important that both parties understand where the boundaries lie for advice relating to the management of current and semi-current records. At an early stage in the relationship, the extent and nature of the 'advice' to be offered by the digital archivist should be defined clearly to avoid potential misunderstandings and liability risks.

Writing guidelines and offering advice

While guidelines written for depositors are of necessity formulaic, the digital curator should avoid alienating potential depositors by being unduly dogmatic or prescriptive. The digital archivist should seek to supplement generic guidelines with ongoing tailored advice which is relevant to the depositor and which accommodates their wishes. See *Appendix B: Guidelines for creators of personal archives*.

Purchasing personal digital archives

In future high value personal digital archives, such as those created by an eminent literary figure or artist, are likely to be sold in auction houses. This raises a number of issues for the collecting institution, including the need to prove authenticity and establish provenance. If the records were held on obsolete media it would also be necessary to prove that the records could be reliably retrieved prior to purchase. Another key consideration would be whether the seller might try and sell copies of the same records to multiple institutions. To prevent such an occurrence a legally binding agreement that guarantees the collecting institution the right to hold the sole complete research copy will be required.

All collecting strategies adopted by a collecting institution must be reviewed regularly in light of technological and cultural changes and with regard to successful strategies adopted elsewhere within the digital archive community.