First Steps in Digital Preservation:
Collections Audit and Business Case

Thom Carter
May 2013
# Contents

## Introduction

2

## Digital collections audit

- Purpose
- Gathering information
- Characterisation tools
- Collection profiles
- Risk assessment
- Tips and resources

3

4

6

6

8

## Business Case

- Purpose
- Preparation
- Background and supporting evidence
- Risks and benefits
- Recommendations
- Next steps: presenting the business case and beyond
- Tips and resources

9

9

10

10

10

11

12

## Appendices

- Business Case: Digital Preservation at Bishopsgate Institute
- Presentation slides
Introduction

Bishopsgate Institute

Founded in 1895, Bishopsgate Institute is home to a historic library and archives. The Institute also hosts cultural events, courses for adults, and a schools and community learning programme. Bishopsgate Library holds special collections relating to the history of London, labour and socialism, freethought and humanism, cooperation, and protest and campaigning. In recent years, the Library has accumulated substantial digital collections, which have mostly accrued through an ongoing project to digitise parts of the existing special collections, and through deposits of born-digital material. As of January 2013, the Library did not have a formal digital preservation strategy in place, nor did it have dedicated storage arrangements for its digital collections.

SPRUCE

SPRUCE (Sustainable PReservation Using Community Engagement) is a JISC-funded project that aims to foster a vibrant and self-supporting community of digital preservation practitioners and developers via a mixture of online interaction and face-to-face events. The events provide support and technical expertise to address specific digital preservation challenges. SPRUCE encourages attendees to apply for funding awards to further develop and embed work from these events at their home institutions. SPRUCE is building on the experience of these activities in order to develop a strong business case for digital preservation, with the aim of supporting and embedding good digital preservation practice beyond the life of the project.

First steps in digital preservation at Bishopsgate Institute

After attending a SPRUCE event in London in July 2012, Bishopsgate Library made a successful application for a funding award for a three-month project that would enable us to take the first steps towards implementing a digital preservation strategy: an audit of the Library’s digital collections, and a business case for the work needed to preserve them.

In addition to carrying out the audit and writing the business case, we also set out to produce this handbook, which documents our work on the project. It describes the processes involved, the tools and resources used, and the lessons learned, and is aimed particularly at those working in smaller institutions in the UK higher education and heritage sectors contemplating taking their first steps in digital preservation. Organisations differ, and our approach may not work for all, but we hope that this handbook will provide a useful starting point for those looking for some practical tips on how to kick start digital preservation activities at their own institution.
Digital collections audit

Purpose

At the outset of the project, the majority of Bishopsgate Library’s digital collections were stored on a network drive shared across Bishopsgate Institute and accessible to all staff. This drive contained material digitised by the Library along with deposits of born-digital records, both of which we consider to be worthy of long-term preservation; it also, however, contained less important information created to support the day-to-day business activities of the Institute, such as digital reproductions produced to fulfil reprographics orders and images for publicity campaigns. In total, this drive contained approximately 100,000 files, comprising 300GB of data: not a particularly large dataset by the standards of some institutions, but certainly large enough to present a challenge.

To preserve the Library’s digital collections, we would need to isolate them from less valuable data and ensure that their location and content were adequately documented. The purpose of the digital collections audit was therefore to gather as much information about the collections as possible. We would use this information to inform preservation planning decisions and to help identify the risks to the collections that would need to be addressed. Moreover, the information gathered for the audit would support our business case for further digital preservation work: it would be much easier to make a case for digital preservation if we were able to accurately articulate the scale and content of our digital collections.

Gathering information

Some of the questions about the collections that we wanted to answer were simple: where were they stored, and where had they come from? Others were more complex: which items within the collections were particularly worthy of long-term preservation, and what specific risks to the collections could we identify based on their technical characteristics?

We made a list of properties that we wanted to gather for each of the collections, and divided these into those that could be gathered automatically using collection characterisation software, and those that would need to be gathered manually by inspecting each of the collections individually.

Properties that would need to be gathered manually were:

- The origin of the collection (we categorised collections according to whether they were created internally, for example as part of an ongoing digitisation programme, or externally and subsequently deposited at Bishopsgate Library).
- The purpose for which the collection was created or stored (we divided the collections into three categories: born-digital archival material, digital surrogates for paper collections, and administrative information created to support the day-to-day business activities of Bishopsgate Institute).
- Whether the collection was complete or still accruing.
- A brief description of the collection’s content.
Properties that could be gathered using characterisation tools were:

- The physical location of the collection.
- When the collection was created.
- The total amount of disk space occupied by the collection.
- The total number of files in the collection.
- A list of the different file formats found in the collection (expressed as MIME types), and the number of files for each.
- If available, we also hoped to extract more detailed technical metadata for files in the collections, such as the colour space for image files, or the page count for PDFs.

We did encounter some difficulties whilst gathering the information for the audit. Manually assessing the origin, purpose and content of the collections was a time consuming process. Moreover, there were inevitably cases that did not fit neatly within the categories we had devised, such as those collections that contained both high-resolution digital surrogates and low-resolution images created for administrative purposes. Similarly, it was hard to decide whether the files in some collections, particularly those which contained images that had been digitised from negatives, should be classified as surrogates or as new born-digital objects; in the end, we chose to group them with other surrogate images created as part of our digitisation programme. Finally, it was not always easy to determine what constituted a ‘collection’. As the directory structure within our shared drive broadly organised files according to their origin and purpose, we decided to adopt a pragmatic approach to this problem: for the purposes of the audit we defined ‘collections’ as top-level directories within this drive.

Characterisation tools

Characterisation tools enable collection owners to extract information about the objects within their collections. This information can then be used to build up a more detailed profile of an institution’s digital holdings and inform digital preservation planning and risk assessments.

There are a number of different characterisation tools available, suitable for a range of use cases. Most are open source and available to download and use for free. A good place to start looking for a tool to fit your needs is the Open Planets Foundation Digital Preservation Tool Registry, which provides descriptions and links for a wide selection of tools.

We used three different toolsets to gather information about our collections for the audit:

1) Bash

Bash is a Unix shell (command-line interface) that is distributed with most Linux systems. It is not a characterisation tool per se, but does provide access to a range of utilities capable of extracting basic characterisation information from digital collections.

We wrote a simple bash script to extract information about the collections (the path, number of objects, a range of modification dates, and total size), which formed the basis of the characterisation profiles we were assembling for each collection.
2) DROID

DROID (Digital Record Object IDentification) is a tool developed by The National Archives. It identifies file formats based on the definitions contained within The National Archives’ file format registry, PRONOM.

We experimented with using DROID to gather information such as file format identifications for the audit. DROID has an easy to use graphical interface, and produces accessible reports summarising the information it has extracted. In the end, however, we chose to use a tool that was capable of extracting a greater range of characterisation information, FITS.

3) FITS

FITS (File Information Tool Set), developed at Harvard University, identifies, validates, and extracts a wide range of technical metadata for various file formats. It wraps several third-party characterisation tools, including DROID, and consolidates their output into a single XML report for each file that it processes. FITS can only be run from the command line.

We used FITS to analyse all of the files in our digital collections. Unlike DROID, FITS does not create reports summarising the metadata it has extracted, so we wrote a few simple Python scripts to process the FITS XML output files and pull out the information that we were interested in across the collections.

We also made use of another tool, C3PO (Clever, Crafty Content Profiling of Objects), which enables users to analyse the characterisation metadata extracted by FITS. C3PO is able to generate graphs and reports based on the information contained in the FITS output files.

Some characterisation tools are easier to set up and start using than others. DROID, for example, provides a graphical user interface and its only prerequisite is that Java is installed. Bash and FITS, by contrast, require users to be comfortable with a command-line interface, and C3PO is dependent on other applications (a database system and a web application framework) that need to be installed before it can be used. Based on our experience with using these tools for our audit, we concluded that the richer metadata extracted by FITS allows for more accurate identification of preservation-worthy data and risks (there were lots of cases, for example, where the reliability of the DROID format identifications were called into question by conflicting results from the tools bundled with FITS). However, to colleagues in a similar organisation looking for a basic characterisation tool, we would still recommend DROID: we think it is the best tool to get up and running and access some meaningful data about digital collections quickly and easily.
Collection profiles

The information gathered manually about the collections was recorded in a spreadsheet. We combined this information with the properties extracted from the FITS metadata to produce a profile for each collection in the following format:

- **Collection:** Howell Archive Photographs
- **Location:** P:/Howell Archive Photographs
- **Dates:** 2005-2012
- **No. of objects:** 116
- **Total size:** 289.10 MB
- **Origin:** Internal
- **Category:** Surrogates
- **Status:** Accruing
- **Description:** Digitised images and documents from George Howell Collection

<table>
<thead>
<tr>
<th>Formats</th>
</tr>
</thead>
<tbody>
<tr>
<td>image/tiff,8</td>
</tr>
<tr>
<td>image/jpeg,107</td>
</tr>
<tr>
<td>application/octet-stream,1</td>
</tr>
</tbody>
</table>

These profiles did not contain the full range of properties extracted by FITS: this information was retained for further analysis later. The primary motivation for creating the profiles was to inform our preservation planning and risk assessment. However, we also intended them to serve as an interim catalogue and collections register for the digital collections.

Risk assessment

Having gathered the required information about our collections, we moved on to the second phase of the audit: a preliminary risk assessment. Several different models for digital preservation risk assessments have been published, some of which are intimidatingly lengthy and detailed. These models vary considerably: some advocate limiting the focus of a risk assessment to threats to the collections themselves, such as hardware or software failure, whilst others suggest that risks to a repository’s parent institution, such as financial insolvency, should also be considered. The models also vary according to how they conceptualise risks: some prefer to focus on the causes behind threats to digital collections, whilst others tend to emphasise their potentially damaging consequences.

For our risk assessment we made use of two of these models: the SPOT (Simple Property-Oriented Threat) model and DRAMBORA (Digital Repository Audit Method Based on Risk Assessment).

SPOT aims to provide a simple model for risk assessment, focused on safeguarding against threats to six properties of digital objects that the authors consider fundamental to their preservation: availability, identity, persistence, renderability, understandability, and authenticity. The model discusses threats in terms of their potential impacts on these properties, providing several example outcomes for each. The authors have also included a useful comparison of other digital preservation threat models. SPOT helped us to determine the different categories of risks that we would need to consider.

DRAMBORA provides a complete online toolkit for a digital repository audit. The toolkit guides users through the audit process, from defining the purpose and scope of the audit to identifying and addressing risks to the repository. Completing a full audit based on the
DRAMBORA model was beyond the scope of our project. However, DRAMBORA does provide a list of over 80 examples of potential risks to digital repositories, framed in terms of possible consequences such as ‘loss of integrity of information’. We used these examples to help build up a list of potential specific risks to our collections.

We also attempted to identify risks using the characterisation information we had gathered for the audit. This process can be challenging, and typically requires some knowledge of preservation risks attached to particular file formats. Nonetheless, we were able to attribute a higher level of risk to those files for which FITS could not provide format identifications, either because the bundled tools did not recognise the file type or because they produced conflicting results. We will need to examine these unknown files more closely so that their format can be determined and documented and any necessary preservation actions can be taken.

We had limited time to allocate to the risk assessment, and our intention for this project was to focus on identifying practical preservation solutions for Bishopsgate Library’s digital collections. We therefore decided to limit the scope of the risk assessment to threats to the collections themselves: we did not include more general risks to Bishopsgate Institute that could indirectly impact the collections. We also decided to frame risks in terms of their causes rather than their consequences: a consequence of a ‘loss of integrity of information’, for example, could be caused by a lack of monitoring of the collections to ensure their integrity. This would help more easily determine the actions that would need to be taken to address these risks, and would enable us to make explicit links between the recommendations made in our business case and the risks we had identified. We grouped risks into three categories relating to different aspects of the management of the digital collections: strategy and staffing, storage, and metadata.
First Steps in Digital Preservation: Collections Audit and Business Case

Tips for getting started with a digital collections audit:

1. Define the purpose of your audit. What will the information gathered be used for?

2. Make a list of the information you will need to gather about your collections, and decide how you will store and present this information.

3. Select a characterisation toolset or devise a manual procedure to fit your information-gathering requirements.

If your audit will include a preliminary risk assessment:

1. Define the scope of the risk assessment. Will you include general risks to your organisation, or limit your focus to your digital collections?

2. Consider how best to conceptualise your risks. A risk assessment focusing on consequences may help to make a more dramatic case for digital preservation, but concentrating on causes will make it easier to justify your recommendations.

3. Select an appropriate risk assessment model to help determine and categorise risks.

Resources:

Open Planets Foundation Digital Preservation Tool Registry
http://wiki.opf-labs.org/display/TR/Home

DROID (Digital Record Object IDentification)

FITS (File Information Tool Set)
https://code.google.com/p/fits/

SPOT (Simple Property-Oriented Threat) model for risk assessment
http://www.dlib.org/dlib/september12/vermaaten/09vermaaten.html

DRAMBORA (Digital Repository Audit Method Based on Risk Assessment)
http://www.repositoryaudit.eu/
Business Case

Purpose

The information gathered for the audit, and the list of risks to the collections that we had identified, formed the basis of a business case for further digital preservation work at Bishopsgate Institute.

A business case is typically used to outline a specific need, and a solution to that need. However, we decided to make a more general case for digital preservation to become a core business activity at Bishopsgate Institute. We would supplement our case with a set of specific recommendations, to be implemented over the next two years, intended to prompt focused action in response to our document. The reasons for this decision were twofold: firstly, our organisation was undergoing a strategic review, and we concluded that this presented us with an opportunity to make the case for the introduction of a permanent new strand to the Library’s activities; secondly, we recognised that we were only in the very early stages of our digital preservation planning, and that more preparation would be required before we could finalise specific plans for future work.

Preparation

The most important factor to consider when compiling the business case was our audience. The document would be aimed primarily at Bishopsgate Institute’s senior management team and, to a lesser extent, the organisation’s governing body of trustees. These two groups would be responsible for making decisions about committing the resources needed to establish a digital preservation strategy. Neither group, however, would be familiar with the challenges of preserving digital collections. The business case would therefore need to provide some explanation, in non-technical language, of these challenges.

With our audience in mind, we selected some key messages that we wanted to communicate throughout the business case. Firstly, as the senior management team takes a holistic view of the organisation, we wanted to stress that a digital preservation strategy would underpin a range of activities at Bishopsgate Institute, including those outside the Library. Secondly, we wanted to emphasise the value of our digital collections: by outlining both how the collections benefit Bishopsgate Institute, and their importance as a historical resource, we hoped to convince our audience of the merits of preserving them.

We used planning exercises developed for use at SPRUCE events to help determine the structure of our business case. These exercises are designed to enable practitioners to identify key stakeholders, skills gaps, and the benefits of introducing a digital preservation strategy for their organisation. We also referred to a new business case checklist created as part of another SPRUCE-funded project at the Institute of Education.

Based on our planning and our consideration of our audience, we determined that the business case should include the following sections:

- Background information about digital preservation and its relevance to Bishopsgate Institute.
- Supporting evidence for the need for a digital preservation strategy, drawn from our audit data.
First Steps in Digital Preservation: Collections Audit and Business Case

- Current risks to our digital collections.
- Benefits of preserving our collections.
- Recommendations for future work.

Background and supporting evidence

We opened the business case with a brief introduction to digital preservation. Our aim was to articulate the seriousness of the problem, to demonstrate its relevance to Bishopsgate Institute, and to summarise the actions that would need to be taken in order to preserve our digital collections. We also pointed to the wider professional context for digital preservation, making reference to similar work at other institutions.

We used the data gathered for our digital collections audit to provide supporting evidence for the need for a digital preservation strategy. We provided basic statistics about the collections, including their total size, location and composition, and the proportion of the data we had audited that we considered worthy of long-term preservation. Most importantly, however, we were able to demonstrate ongoing trends in the development of the collections, including their origins and growth. We included a graph to highlight the acceleration in the rate of growth of the collections over time, based on the file modification times extracted by FITS.

Risks and benefits

The list of risks to our digital collections included in the business case was taken directly from our risk assessment. We provided a brief explanation for each risk, using terms that would be familiar to our audience: the lack of technical and descriptive metadata for the collections, for example, was categorised under the general heading ‘cataloguing’. We also sought to demonstrate the possible impact of risks beyond the collections themselves, such as the potential harm to Bishopsgate Institute’s reputation should archival material be lost or damaged.

The SPRUCE stakeholder analysis exercise had shown that there were would be a range of potential stakeholders in a future digital preservation strategy at Bishopsgate Institute, including the Library, other departments within the organisation, and, most importantly, our users. We therefore needed to demonstrate that introducing this strategy would yield benefits for all of these groups. The positive outcomes that we listed included both those that would be advantageous for the Library, such as the capacity to accept deposits of new digital collections, and those that would be of wider benefit to Bishopsgate Institute, such as income generation. We had previously completed a SPRUCE exercise intended to help strengthen a case for digital preservation by linking potential benefits to existing institutional objectives. We consulted a number of strategy documents, including the Library strategy and Bishopsgate Institute’s digital strategy, and added quotations from these documents to the business case. This enabled us to make explicit links between the benefits we envisaged and our organisation’s existing strategic objectives.

Recommendations

Developing and implementing a full digital preservation plan for Bishopsgate Library was beyond the scope of this project. However, we were in a position to make a series of recommendations, based on the risks we had identified, which would help the Library to
achieve this goal. We outlined a roadmap of actions to be carried out over the next two years, addressing risks in the three key areas of staffing and strategy, storage, and metadata.

The roadmap included some tasks, such as developing a digital preservation plan and preparing for the introduction of a Digital Asset Management System, which we knew would require substantial further work. We recognised that if we hoped to implement a successful and sustainable digital preservation strategy, we would first need to secure adequate staff resources for this. Our principal recommendation, therefore, was for a new permanent position of Digital Archivist to be created in the Library, with the postholder being responsible for the development and preservation of the Library’s digital collections.

Next steps: presenting the business case and beyond

Once the business case was completed, we presented it to our senior management team. Our aim was to make the strongest case possible for the recommendations set out in the document. The presentation included a brief introduction to the work carried out for the project, and a description of the Library’s digital collections, complete with visual examples. We focused on articulating the problem of digital preservation, including the particular risks to the Library’s collections, again with visual examples such as a damaged image file discovered during our audit. Finally, we sought to emphasise the extent to which a digital preservation strategy would underpin other activities at Bishopsgate Institute, both now and in the future.

We were aware that a decision about whether to commit the necessary resources for the creation of a digital preservation strategy would not be made immediately. In order to sustain the momentum we had built up whilst working on the project, we identified a series of smaller objectives that we could pursue in the meantime. These would help to embed the work we had already carried out, and would substantially improve the condition of our digital collections:

- Liaise with our IT technician to arrange for a dedicated storage drive for the Library to be set up. Once this was done, begin copying collections classified in the audit as born-digital or surrogates to this drive.
- Use data gathered in the audit to create a register of the digital collections, including basic descriptions and physical locations.
- Create a verifiable manifest for each collection, again using data gathered in the audit. Design and implement a procedure for monitoring the integrity of the collections, by checking them against these manifests.
- Design and implement a procedure for ingesting digital collections, in preparation for new digital archive deposits.
First Steps in Digital Preservation: Collections Audit and Business Case

Tips for getting started with a digital preservation business case:

1. Be clear about what you would like the business case to achieve. This will form the basis of the recommendations you make in your document.

2. Consider your audience. How should the content or language of the document be tailored to suit your readers?

3. Research your organisation’s strategic objectives. Ensure that your recommendations are aligned as closely as possible with existing priorities.

4. Assemble your evidence. Support your case by demonstrating the current condition of your collections, the risks that they face, and the benefits of preserving them.

Resources:

SPRUCE business case resources

Business case checklist
http://wiki.opf-labs.org/display/SPR/Checklist+for+writing+a+business+case+for+digital+preservation
Business Case:
Digital Preservation at Bishopsgate Institute

Thom Carter
March 2013
Executive summary

Bishopsgate Library holds substantial digital collections. Some of these collections have been created internally as part of the Library’s digitisation programme, whilst others have been received from external depositors. The digital collections have grown steadily over the past five years, and will continue to grow in the years to come.

The Library’s digital collections are an important historical resource, and a valuable asset to Bishopsgate Institute. Digital collections are more vulnerable to loss or damage than their paper counterparts. The Library must therefore take action to ensure that its digital holdings are preserved according to recognised standards.

In October 2012, the Library made a successful application for a funding award from SPRUCE, a JISC-funded project that supports digital preservation activities in UK institutions. The award was intended to enable the Library to take the first steps towards implementing a digital preservation strategy: an audit of the digital collections, and this business case for further work.

This business case outlines the current risks to the Library’s digital collections and highlights the benefits of preserving them. It concludes with a set of recommendations which should be implemented over the next two years. These include reviewing Library staffing provision to ensure that sufficient resources are allocated to work on the digital collections, and a series of specific measures relating to the management and storage of these assets.
Contents

1. Background 3
2. Digital collections audit 4
3. Current risks to the digital collections 6
4. Benefits of preserving the digital collections 8
5. Recommendations 10
1. **Background**

Over the last ten years, Bishopsgate Library has collected a substantial quantity of digital material. This has mostly accrued through an ongoing project to digitise parts of the Library’s collections, including the Morning Star Photographic Archive and the Feminist Library Pamphlet Collection, and through deposits of digital records from organisations such as the British Humanist Association and History Workshop. We expect the Library’s digital collections to grow still further in the coming years, as digitisation work continues and an increasing number of archive deposits are received in digital form.

In a recent report on archival practice in the United Kingdom, *Archives for the 21st century in action: refreshed 2012-2015*, The National Archives recommended that over the next three years archives should ‘ensure clear strategies are in place to enable the selection, management and preservation of digital records’. Digital collections are more vulnerable than their paper counterparts: they are susceptible to a range of threats, including system failures, human error and technological obsolescence, which can result in the loss of the files they contain, or leave those files corrupted or unreadable. Bishopsgate Library does not yet have the necessary policies and procedures in place to safeguard against these threats and ensure the long-term preservation of its digital collections.

Organisations such as The National Archives, the British Library and the Digital Preservation Coalition have collaborated to develop standards and best practices for the care of digital archive collections. Bishopsgate Library must implement a preservation strategy based on these standards for its digital collections, just as we have for our paper collections. This will involve establishing procedures for accessioning, storing and monitoring digital collections. Digital preservation is a continuous process, which will need to be adequately resourced with both staff time and IT provision if it is to be successful and sustainable.

This business case describes the current state of the Library’s digital collections and outlines the risks that they face. It highlights the benefits of preserving these collections, and sets out a series of recommended actions that will need to be taken to ensure that these benefits can be realised.
2. Digital collections audit

In October 2012, the Library made a successful application for a funding award from SPRUCE (Sustainable PReservation Using Community Engagement), a JISC-funded project that provides support for digital preservation activities in institutions throughout the UK. This funding award has enabled the Library to take the first steps towards implementing a digital preservation strategy: an audit of the Library’s digital collections, with the aim of evaluating their scale and content, and a business case for further digital preservation work.

The audit confirmed that the majority of the digital collections are currently stored on the ‘P’ drive, a network drive shared across Bishopsgate Institute. A small number of collections are stored in the Library’s directory on the ‘N’ drive. In total, the collections include nearly 100,000 files, amounting to over 300GB in size. The collections have grown steadily over the past ten years, with the rate of growth accelerating markedly in the last three years (Fig. 1).

![Fig. 1: Increase in number of files in digital collections by year](image)

The collections have been classified into three categories:

- administrative (files created to support the day-to-day business activities of Bishopsgate Institute)
- surrogates (digitised reproductions of material from the Library’s paper collections)
- born-digital (archival collections received by the Library in digital form)

Collections falling into the latter two categories, which account for 56% of the Library’s digital holdings, are particularly worthy of long-term preservation. The creation of the digitised surrogates has required a significant investment of staff time, and these resources both enhance access to the collections and provide a source of income through the Library’s reprographics services. Like our paper collections, the born-digital collections contain unique archival documents, and should be preserved as a valuable research resource.
The audit indicated that the majority of the files in the collections (over 80%) were images, with the remainder being documents, videos and audio files. Most have been created within Bishopsgate Institute, many as part of the Library’s ongoing digitisation programme, but a small number of born-digital collections were created externally and deposited at the Library in digital form. This number will rise as regular depositors of archival material switch from paper to digital deposits, and as the Library establishes new digital initiatives such as the planned project to create an archive of material documenting the Occupy movement. These born-digital collections often include a range of different types of digital content, and will therefore present more complex preservation challenges than the Library’s digitised image collections.
3. Current risks to the digital collections

The current arrangements for the administration and storage of the Library’s digital collections expose them to a number of risks. If left unaddressed, these risks could impact both the collections themselves and Bishopsgate Institute’s reputation and future funding prospects. Should digital collections be lost or damaged, the Library’s status as a trustworthy repository for such resources and its ability to attract both digital archive deposits and funding for digitisation projects would be severely diminished.

The Library has identified the following risks to its digital collections, which should be addressed as soon as possible:

Strategy and staffing

- **The Library has no digital preservation plan in place.** No established procedures exist for accessioning new digital collections or for storing and monitoring existing collections.

- **Staff do not have the time or expertise required to manage the Library’s digital collections effectively.** There is currently no provision in Library staff job descriptions for the management of the Library’s digital resources, and most staff have not received adequate training in this area. Most of the work on the Library’s digital collections is currently performed by the full-time Library and Archives Assistant, in addition to his contractual responsibilities. This is not a sustainable arrangement, as successors to his post could not be expected to continue this work.

Storage

- **There is no dedicated storage system for the Library’s digital collections.** The digital collections are stored on a shared network drive, leaving them open to unauthorised access and modification.

- **Digital collections are not monitored to ensure their integrity.** Library staff are unable to determine whether or not digital collections have been modified or damaged since they were created or acquired.

- **Digital collections are stored on ageing physical media.** Some digital resources are currently stored on floppy disks and other media, which are vulnerable to obsolescence and decay.

- **Backup arrangements for the digital collections are inadequate.** Insufficient copies of the Library’s digital resources exist to safeguard against system failures and ensure long-term preservation of the collections.

Cataloguing

- **There is no register of the Library’s digital collections.** Basic descriptions and physical locations of new and existing digital collections are not recorded.
Business Case: Digital Preservation at Bishopsgate Institute

- **Digital collections are not catalogued.** Most of the Library’s digital resources are not discoverable by Library users or staff and no supplementary administrative or contextual information is available.

- **Technical information about the digital collections is incomplete.** Details about the Library’s digital resources that are required for effective preservation, such as file format identifications, are not available.
4. Benefits of preserving the digital collections

The Library’s digital collections are a valuable asset for Bishopsgate Institute. The importance of the collections will increase over time as their size as a proportion of the Library’s total holdings grows.

Taking appropriate measures now to ensure the preservation of the collections will yield benefits not only for the Library service but for Bishopsgate Institute as a whole, particularly in the areas of profile-raising and income generation. These benefits are closely aligned with the objectives set out in the *Library and Archives Action Plan* (April 2011) and the Institute’s *Digital Strategy* (January 2012).

1. Preservation of the Library’s digital collections

Digital collections will be preserved according to recognised standards, as is the case with the Library’s paper holdings.

*Provide and maintain adequate and suitable accommodation for the various collections, meeting recognised standards for archival storage (Library and Archives Action Plan – April 2011)*

2. Increased access for wider audiences

Preservation of the Library’s digital collections will ensure that they can continue to be accessed both locally and remotely by a wide audience, including students and academic researchers. Several of the digital collections are already viewable on the Library’s online catalogue and web pages, and we intend to add more in the coming months.

*Enact a comprehensive programme to digitise items from the Library’s archival collections and present through Archives Online on the Institute’s website (Library and Archives Action Plan – April 2011)*

3. Knowledge and expertise

Staff will develop knowledge of digital preservation challenges and solutions.

*Ensure sufficient and suitably trained staff to provide services (Library and Archives Action Plan – April 2011)*

*Ensure that new and appropriate technological developments are embraced and resourced (Digital Strategy – rev. January 2012)*

4. Digital acquisitions

The Library will attract new deposits of digital archive collections and will offer assurances that these collections will be curated and preserved according to recognised standards.
Use the Collection Development Policy to inform ongoing acquisitions and disposals in the context of national uniqueness and profile-raising (Library and Archives Action Plan – April 2011)

5. Opportunities to participate in external projects

The Library, along with the Schools and Community Learning team, will continue to participate in partnership projects to create online resources drawing on digital archive collections. In addition to the partnership with SPRUCE for this project, recent collaborations have included the Exploring 20th Century London website, an online resource for History Workshop Journal, and On the Record’s Sounds from the Park project.

Seek funding for additional project staff (Library and Archives Action Plan – April 2011)

Utilise suitable opportunities to work collaboratively with other relevant institutions and on external projects which will enhance access to and support for our programmes (Digital Strategy – rev. January 2012)

6. Participation in the wider digital preservation community

The Library will work with other institutions to share digital preservation challenges and solutions.

Join and contribute to development and activities of relevant professional networks (Library and Archives Action Plan – April 2011)

7. Income generation

The Library’s digital collections provide a source of income for the Institute (through charging for use of images in commercial publications, for example). Preserving these collections will protect this revenue stream and allow us to explore opportunities to increase this income, such as the introduction of an online image database.

Explore income generating opportunities e.g. through increased services to users, publishing etc. (Library and Archives Action Plan – April 2011)

8. Organisational record-keeping

Bishopsgate Institute will ensure that it meets its legal obligations for record-keeping, and that its digital records are kept according to the same high standards as is paper records.
5. Recommendations

Bishopsgate Library has functioned as a historical repository for over 100 years, and recent years have witnessed a dramatic professionalisation of the Library service. If the Library is to maintain these standards, and remain relevant as a memory institution, it must respond to the challenges of the changing landscape of archival practice in the digital age. Given the Library’s focus on and reputation for collecting contemporary material, especially that relating to protest and campaigning, these challenges are both pertinent and pressing.

The Library therefore aims to put in place a sustainable strategy to mitigate the risks to its digital collections and ensure their preservation. To achieve these objectives, we recommend that the following actions be taken:

**Immediately**

- Library staffing provision should be reviewed. Current Library staff resources are not sufficient to address the risks and implement the recommendations set out in this business case. The Library therefore recommends that a new permanent position of **Digital Archivist** be created, with the following responsibilities:

  1) Ensure the preservation of the Library’s digital collections, by implementing the recommendations set out in this business case.
  2) Develop the digital collections by continuing the digitisation programme, acquiring and cataloguing deposits of digital archive material, and pioneering new initiatives such as the current collaboration with Occupy.
  3) Disseminate the digital collections via the Library’s online catalogue and other outlets such as Culture Grid.

Suitable candidates for this role would possess or be working towards a relevant professional qualification, and be able to demonstrate a good balance of technical and information management skills. The creation of this position will enable the Library to continue to grow its digital collections, whilst ensuring that they are well-used and properly cared for.

- Collections classified in the audit as born-digital or surrogates should be moved to a dedicated storage drive, with limited access permissions. The backup arrangements for this data should be reviewed to ensure that they are sufficiently robust.

- A procedure to monitor and ensure the integrity of the digital collections should be established.

- A permanent collections register, including basic descriptions and physical locations of all the digital collections, should be created using the information gathered during the audit. New digital collections should be added to the register as they are acquired.
Business Case: Digital Preservation at Bishopsgate Institute

Within six months

- A formal digital preservation plan should be drawn up, based on the findings of the collections audit. This should include policies and procedures for the ingest, storage and monitoring of the digital collections.

- Collections stored on ageing physical media should be moved to the Library’s dedicated storage drive.

Within one year

- The Library should determine hardware and software requirements for a digital asset management system, to facilitate storage of and access to its digital collections.

- Technical and descriptive cataloguing of the existing collections should be completed.

Within two years

- The Library should deploy a digital asset management system, and import existing collections.

- To evaluate progress and identify areas for further work, the Library should assess its digital preservation policies and procedures using a recognised accreditation standard for trusted digital repositories such as the Data Seal of Approval.
Digital Preservation at Bishopsgate Institute

What is digital preservation?
• The actions required to ensure that digital records remain accessible over a long-term period.
• “Archives should ensure clear strategies are in place to enable the selection, management and preservation of digital records.”
  The National Archives, Archives for the 21st Century in action: refreshed 2012-2015

SPRUCE Project
• Funded by SPRUCE/Jisc
• First steps in digital preservation
• Two main elements:
  - an audit of our digital collections
  - a business case for further work

Digital Collections
• ~100,000 files
• Majority worthy of long-term preservation
• 90% created internally
• Mostly images
• Growing!

Increase in the number of files added to the digital collections by year
Risks

**Storage**
- Lack of dedicated storage
- Collections not monitored
- Collections stored on ageing removable media
- Backups

**Cataloguing**
- No collections register
- Collections not catalogued
- Incomplete technical information

**Strategy and staffing**
- No digital preservation plan
- Insufficient staff resources

Damage to the collections → Loss of access to the digital collections → Technical obsolescence → Lack of contextual information
Recommendations

- Six months
  - Formal digital preservation plan drawn up
  - Collections moved to dedicated storage
  - Procedure for monitoring collections established
  - Permanent collections register established

- One year
  - DAMS requirements analysis
  - Technical and descriptive cataloguing

- Two years
  - Deploy DAMS
  - Review and evaluate

Why should we do this?

Digital Preservation

Collection Development
Electronic Records Management
Projects and Partnerships
Image Database
Audience Development

Questions?