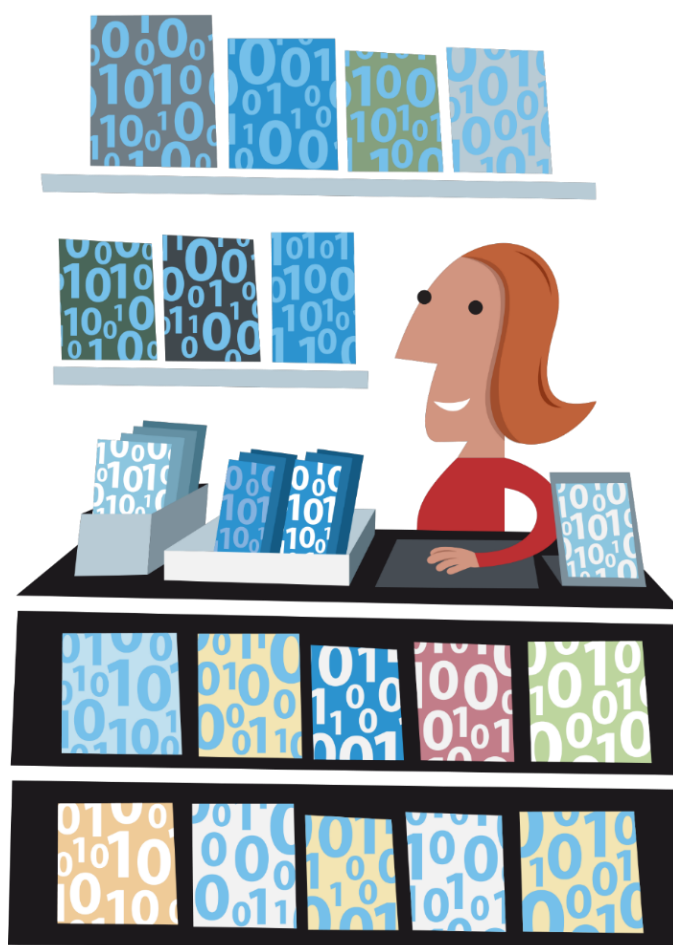


DIGITAL PRESERVATION FOR SMALL BUSINESSES:

AN INTRODUCTORY GUIDE



Why is digital preservation important for businesses?



Digital information plays a central role in most aspects of our lives. Running a small business is no exception. Important information that would once have been produced on paper is now created, shared and maintained as digital information (for example, annual reports, minutes of meetings, contracts and other legal agreements, publicity or marketing materials, financial returns and correspondence). Small businesses create all these types of digital information and more, and need to ensure they can be accessed for as long as necessary.

Information may need to be kept for many reasons. You may be required to maintain certain types of information to comply with regulations, some information may provide evidence of decisions or transactions made, other information may simply be valuable in recording the history of your business. Ensuring you can access and use digital

information presents very different challenges compared to ensuring ongoing access to more traditional physical resources.

If you have a handwritten document, drawing or a printed book, you will be reasonably confident that you can store it and reuse it at a later date. Physical items such as these are generally quite robust, and will even hold up well to being put in a cupboard and forgotten about. It would be reasonable to expect that you will be able to go back years later and access, understand and use these materials.






However, digital materials are more complex and require more proactive management. Take, for example, a text document saved on a rewritable CD. To be able to access and use this document you must first have a computer with a CD drive, and a compatible piece of software installed that will allow you to open the document and display it on screen correctly. Computer hardware and software are constantly changing, and newer versions are not always fully compatible with earlier products. You also need to consider the longevity of the media on which your content is stored. Rewritable CDs are surprisingly fragile and susceptible to data loss, so there is no guarantee that your document will still be accessible at all. If your CD was put in a cupboard and forgotten about, we would have much lower confidence in being able to retrieve and use the information in the future.

Digital preservation is a term used to describe the activities that are carried out to mitigate the risks to digital content over the long term. It is ultimately about enabling access to digital content, both now and at some unspecified point in the future. Though digital preservation might seem quite a specialist activity, the advice contained within this guide focuses on common sense management of digital information and highlights actions you can incorporate into your day-to-day working practices in order

to have greater confidence in your ability to manage and preserve the valuable information that you hold.

Digital preservation myths

There are several common misconceptions about digital preservation:

-  **Digital preservation is the same as digitization** – Digitization is the process of turning physical materials such as documents or photographs into digital content using a scanner or digital camera. This process generates more digital content that needs to be properly managed and sometimes preserved for the future, but it doesn't guarantee the preservation of that content.
-  **Digital preservation is just about backing-up your digital content** – Whilst having a backup copy of your digital content is an important part of managing and protecting that content for the future, it isn't the only thing that you need to consider to ensure that information is safe.
-  **Digital preservation is about keeping hold of everything forever** – Businesses may create substantial quantities of digital information as part of their day-to-day work, but it is unlikely that all of it has long-term value. Some information will only need to be kept for a specific period of time and can then be deleted. When managing digital content it is important to understand what needs to be kept and for how long, and to ensure that content that is no longer of value is deleted. Clearing out unwanted content periodically makes it easier to identify and manage the more valuable digital materials that need to be kept for the longer term.
-  **Digital preservation is a one-off activity** – Digital preservation requires a series of ongoing actions over time to ensure that digital content can continue to be read and understood. Your digital content will need to be revisited many times so you can carry out further checks and preservation actions. In short, do not believe anyone who tells you that they have completed the task of preserving your digital content!
-  **We could just print everything out instead...** – There may be situations where having a hard copy print-out of a digital document would be useful as a 'backup' to the digital copy, especially where an organization does not have confidence in its ability to manage and maintain digital content. However, printing out all digital content with long-term value is unlikely to be the best strategy. A hard copy of a digital document will not retain the functionality of the original digital content and cannot be used in the same way. For example, a print-out of a spreadsheet of sales figures cannot be analyzed and interrogated in the same way as the original digital version.

Spotlight on digitization

Digitization is a strategy frequently carried out by organizations to enable easier access to, and sharing of, materials they have in physical form. Typically, this involves using a scanner or digital camera to create digital copies of physical photographs, correspondence or reports. As noted in the 'myth busting' section, digitization is not digital preservation, but it is helpful to ensure that the digital copies you are creating are fit for purpose for as long as needed. Here are some points to consider when digitizing physical materials.

- ✓ Give thought to the file format you will use for storing the digitized content. An image format such as JPEG or TIFF should be selected for visual materials such as photographs or diagrams, whereas a format more suitable for storing text, such as PDF/A, may be more appropriate for documents and reports.
- ✓ Ensure the digital copy you are creating is high enough quality for the intended purpose. For example, a photograph will need to be scanned at a higher resolution if you are intending to use it in a promotional brochure or poster.
- ✓ When digitizing multiple items, stop periodically and check that the results are what you are expecting.
- ✓ Manage the digital copies well, following the guidelines in this introductory guide in order to maintain them for as long as required.
- ✓ Retain the original physical materials if they are of value to you. The digital copy should not be seen as a replacement for the physical original unless you have absolute confidence in both the digitization process and your ability to manage and preserve the resulting digital files.



Ten steps towards digital preservation

This section provides guidance to help you keep your digital content safe for as long as necessary. It doesn't provide a full solution to digital preservation, but does offer some small and manageable steps that you can take to give you greater confidence in your ability to maintain digital content, enabling it to be handed over to the next generation, or perhaps even a digital archive or repository at some point in

the future. Some of the steps mentioned in this guide relate to good records management. Records management refers to the ways information is managed within an organization throughout its life cycle, from the time of creation to the time it is either archived for the longer term or no longer required.

1. Think about the future now

Few people enjoy the task of sorting out a mass of digital documents that have been generated over the years with little thought for their future reuse. Avoid this eventuality by managing your digital content well *from the start*. From the point that you create digital information, there are several things you can do to make it easier to look after in the future, and some of these are described in the steps below.

2. Decide what to keep

It is important to understand *which* digital content has value for the longer term. Some content that is created within a business will be useful only for a short period of time. There may be regulations that require certain types of information to be kept for a specific period, after which it can be deleted. There may be efficiency savings or financial gains to be had from information reuse over an extended time frame. Some information may need to be kept to retain corporate memory. Content documenting events of historical interest may need to be kept indefinitely. Knowing which information has value, and for how long, will help you understand where best to target your efforts and make good decisions around its management.



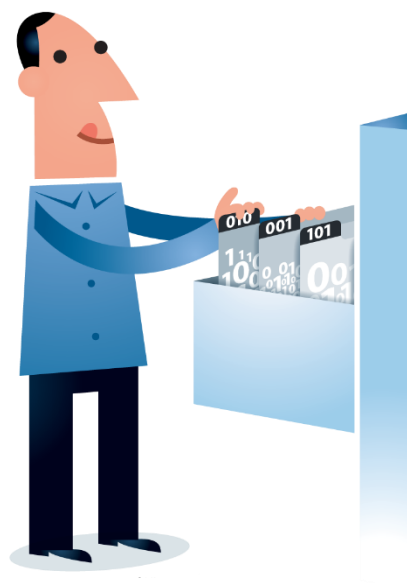
Perhaps surprisingly, deleting content that is no longer required can be an important digital preservation activity; it not only frees up storage space, but enables the more valuable content to be more effectively located and managed. Note also that you may be legally required to delete some information after a specific time in order to comply with data protection legislation.

3. Name your files well

When saving your digital files for the first time, ensure you give them descriptive and meaningful file names that will help you (and others) understand their purpose. Some organizations create internal rules or guidance for file-naming so there is consistency where multiple members of staff are creating similar types of digital content. If you have no such guidance in place, here are some simple rules to get you started:

- ✓ **Use a clear and concise filename that describes what the content is.** This is the most useful thing you can do to ensure that you can search for, and find, the digital content again when you need it. In an ideal world you shouldn't have to open a file to find out what it is!
- ✓ **Use a date in your file names where appropriate.** The date can be an important piece of information so do use it within your file name if this will help to provide context to the digital content. If it is important to be able to order your files chronologically, make sure the year is the first element of the file name, for example, 2022-05-01_sales_figures.xlsx.
- ✓ **Choose a consistent method for recording the file version.** Where multiple versions of the same digital content are created, it is important to be able to locate the latest version. Information about version can be added to a filename (usually at the end) to provide greater clarity, for example, 2022_annual_report_draft01.xlsx.
- ✓ **Use alphanumeric characters (a–z, 0–9), dashes and underscores** to create your filenames and try not to use spaces or special characters such as !* & %. Use of non-standard characters in filenames can create challenges with future management of the files.

4. Use folders to organize your files



Make use of folders (also called directories) to store your digital content in an ordered fashion. Hierarchies of folders can be created to store digital content in meaningful and logical groupings based on their function and purpose.

Folders should be carefully named to provide context about the content within them. Guidance for folder naming is very similar to that described for file naming (above), but most importantly, folder names should be clear, concise and descriptive. Anyone browsing the folder structure should be able to gain some understanding of what is in a folder from its name before opening it. This is helpful for both current and future users.

Spotlight on emails

Emails are a valuable digital record, providing important evidence of decision making and key transactions relating to business activities; however, they are also challenging to maintain and manage correctly and are often neglected. Staff working in an organization may have very different approaches to managing email, perhaps deleting too many or letting ephemeral or junk messages pile up, making it hard to find those of value.

The most important thing you can do to facilitate the preservation of emails relating to your business is to ensure you understand which emails are of importance longer term and put in place some basic measures to protect them:



- ✓ Does your business have a policy or procedure on managing emails? Is there an understanding of which emails have long-term value? If this doesn't exist, consider creating some guidance to be followed by all staff.
- ✓ Establish a methodology for flagging up messages which need to be kept for the long term. Folders could be used to file important emails, keeping them in one place. Some systems may also allow you to 'tag' emails which can help you (and others) understand the context. An 'archive' folder or tag could be used to identify those messages that you want to keep for the long term.¹
- ✓ Equally important as knowing which emails to keep is knowing which emails to delete. If you don't regularly delete emails your inbox will very quickly become unmanageable. Ensure all staff within the business know which emails should be deleted as well as which should be retained.
- ✓ Remember that your sent messages may be as important as the messages in your inbox in providing a record of your business activities. Find out whether sent messages are retained within your email system and establish a means of flagging up which are important.
- ✓ Individual messages can be saved outside your email system where necessary by using the functionality available in your email software; for example, 'Save as...' or 'Download as...'. The 'Print' option may also be used to print an individual email to a PDF file. Any attachments, linked content and other emails in the thread may need to be captured and saved separately.
- ✓ Create a backup of your email on a regular basis. The email software or service that you use should provide a backup or export option.
- ✓ Ensure your email system isn't set up in a way that could inhibit the preservation of your emails. Does it automatically delete emails after a specific time period? Is there enough storage capacity for the volume of email received?
- ✓ Keep business and personal emails separate. This will help with the long-term management of your email.

¹ Don't be fooled by some email systems that claim to 'archive' or 'auto-archive' your email for you. For the most part, they are just moving old emails out of your inbox to get them out of your way – this is not the same as you selecting emails of value and consciously moving them to an 'archive' folder!

5. Document your files

Consider whether your digital content requires any additional documentation to enable yourself or others to understand it, either now or in the future. For some digital content, a good descriptive file name and logical placement within a folder structure can provide all the context that is required, but in some cases additional information is needed.

Here are some examples of documentation that may be required for your digital content:

- **A document, report or promotional leaflet** may be hard to interpret in the future if it doesn't include key pieces of information such as author, version or date of creation or publication. Ensure this information is recorded within the document itself or as part of the file or folder name where possible.
- **A set of images** taken on a digital camera or scanned from physical photographs may be hard to interpret without accompanying information detailing date taken, location, and perhaps even the names of people photographed. This can be a particular challenge where photographs retain default file names allocated by the camera or scanning software. Consider renaming images with descriptive file names that capture key information, or create documentation in a spreadsheet or accompanying document that provides the necessary descriptive information to enable searching and retrieval of the images.
- **A spreadsheet or database** may include ambiguous column headings or use abbreviations or acronyms that need further explanation. Ensure that column headings are descriptive and can be easily understood. Fuller documentation may be created and provided within a separate document where necessary. This may be more important for a complex database, for which the relationships between individual tables will also need to be recorded.



6. Choose a file format likely to last

A file format is a standard way that information is stored on a computer. It tells you (and the computer) which software package should be used to open and view the content. A huge number of file formats are available, some very widely used (for example, the Microsoft Word document) and some much more specialist (for example, the native file format produced by accounting software which can only be opened within that application). Different file formats have different risks associated with them, and some are expected to have greater longevity than others.



Technology changes rapidly, and this causes problems when managing digital content. As new generations of software phase out support for older formats, it can become a challenge to find the software to open files that were created several years previously.

Which file formats will last into the future? Even the digital preservation experts don't really know the answer to this question! In fact, it might be safer to assume that no file format will be readable forever. Those organizations charged

with preserving digital content for the long term will often have plans to update file formats over time to ensure content can still be read with modern software, and this action may need to happen many times over the life of the digital content as technologies change.

There are, however, several points to consider when deciding which file formats your business should use.

- ✓ **Simple file formats** that can be opened in a basic text editor (such as Windows Notepad) are likely to be easier to open and read in the future. Examples of simple file formats include plain text files (TXT) and text-delimited formats such as CSV (comma-separated values).
- ✓ **File formats that are widely used** are often considered to be lower risk than more obscure or niche formats. A critical mass of users helps provide a reason for software providers to continue to provide support for those formats (for example, through backwards compatibility or import routines). Examples of file formats that are widely used include Microsoft Word, Microsoft Excel and the JPEG image format.
- ✓ **File formats designed for archiving** have been specifically designed with longevity in mind and give us greater confidence that we will be able to access the content in the future. If you are creating PDF files, consider saving your files in the archival version of the PDF standard (PDF/A). Standard Microsoft Office applications do allow you to save as PDF/A but the options to do this are well hidden.²

² Go to 'Save as...' and select 'PDF' as your target format. Click through to additional options where you can specify exactly how you want your PDF file to be created. Select 'PDF/A compliant' to ensure your PDF file conforms to the archival PDF standard.

Spotlight on file format conversion

Do you have digital content in older file formats that you are concerned you might not be able to access? If so, consider taking action to keep the information accessible. File format conversion is the process of converting a file to a more suitable format for accessing both now and into the future.

File format conversion sometimes happens automatically. For example, when you open an old Microsoft Word document in a newer version of that software, it may prompt you to update your document to the latest version. If you agree to this, the software will do this format conversion for you. In other situations, you may need to be more proactive and initiate the file format conversion yourself. Perhaps you have an old spreadsheet of financial information that you would like to access, but you no longer have the original software it was created in? You may find you can open this file or import it into a more recent software application such as Microsoft Excel. If you can view the data in Excel, you can use the 'Save as...' option and select a different file format, for example, the CSV (comma-separated values) format or a modern Excel spreadsheet format.

However the file format conversion is carried out, do take care to check that the digital content is as you expect and that nothing has been lost in the conversion process. Since this process is usually irreversible, consider keeping a copy of your original file in case something goes wrong. The types of check you might carry out will depend very much on the content you are working with:

- Does a converted document have the same number of pages as the original? Does it have the same number of words? Are embedded elements (for example, images or tables) still present?
- Does a converted spreadsheet have the same number of rows and columns of data? Are the original formulae present within the converted file, and do they still work?
- Does a converted image look the same as the original, even when you zoom in?

7. Storing your files

Storing your digital content safely and securely is one of the most important things you can do to protect valuable content for the long term. The tips below will help you to store your content well and mitigate some of the most common risks encountered with storage.

- ✓ Understand where your digital content is currently stored. Storage solutions used within a business often grow organically over time in response to specific needs. Perhaps you have content scattered across computer hard drives, stored in the cloud (for example, Google Drive, Microsoft OneDrive or Dropbox), or on portable media (such as portable hard drives, memory sticks, CD/DVDs). If staff use their own devices for work purposes, some content may also be stored on personal laptops



and mobile phones. Getting a clear picture of where your valuable content is located is the first step in protecting it.

- ✓ Consider consolidating content into one primary location where you can easily manage it. A shared or networked drive space or a cloud-based storage solution may be most appropriate.
- ✓ Use a reputable backup application to copy data to one or more external hard drives on a regular basis. The frequency of backup will depend on the value of your content and its rate of change. Ensure you disconnect the drives when not in use. Adding an additional cloud-based storage solution will provide further resilience and can be configured to automatically make copies of your primary digital content store.
- ✓ Media such as tapes or hard disks have a finite lifetime and can fail with little warning. Periodically replace old hardware with new, and carefully copy your digital content, carrying out checks to ensure it has been successfully transferred. Tools are available that can help you to copy content and verify this was carried out successfully (for example, TeraCopy).
- ✓ Ensure any cloud-based storage accounts are not tied to a single member of staff. Should they leave your company, the account might be closed and your files lost. Use a corporate account, and put in place good documentation, clear management policies and agree responsibilities for its maintenance.
- ✓ Consider who has the ability to edit, delete or overwrite your content. Human error, neglect or accidental deletion are the biggest risks to your content so guard against these by ensuring access rights are appropriately managed.
- ✓ Natural disasters, such as fire or flood, have the potential to destroy all data held on a site. Keeping several copies of your content in different physical locations (or in the cloud) will significantly reduce risks.

8. Maintain your computers

Ensuring that all computers and laptops used in the course of your business are kept secure with strong passwords and up to date with security patches and anti-virus software is an important way of mitigating external threats to your digital content. The *Small Business Guide on Cyber Security* from the National Cyber Security Centre in the UK provides helpful, practical advice on how to protect your digital content: <https://www.ncsc.gov.uk/collection/small-business-guide>

9. Consider a longer-term solution for your digital content

The advice within this document will help you to look after your digital content well enough so that you and other staff working with you can continue to access and use it, but it does not provide a complete solution to digital preservation. Perhaps you have digital content on old physical media formats (for example, floppy disks, CDs or DVDs) that you can no longer access? Perhaps some of your content is in file formats that you no longer have the software to open? Perhaps you are simply struggling to

effectively manage increasing volumes of valuable digital content, or are keen that it is made more widely available?



You may wish to consider depositing or donating content relating to your business activities to an archive (this could apply to both physical and digital content). An archive will be able to manage and preserve your files effectively for the future and will also enable access to the information as appropriate.

Do contact your local archive if you would like to discuss this with them. They may also be able to provide you with more detailed advice to supplement the guidance in this document.

Another option available to you would be to preserve your digital content yourself using software that has been specifically designed for this purpose. There are a number of digital preservation systems available, some of which offer free entry-level solutions or are freely available as open-source software. Digital preservation systems typically include functionality to deposit digital content, describe content, carry out file format conversions, and additionally may provide storage and enable discovery and access.

10. Find out more

If you are interested in learning more about digital preservation lots of information is available online, but the following sources from the Digital Preservation Coalition are particularly recommended:

- **The Digital Preservation Handbook**
(<https://www.dpconline.org/handbook>). An online resource providing a thorough introduction and overview to the key areas of digital preservation.
- **Digital Preservation Technology Watch Guidance Notes and Topical Notes Series**
(<https://www.dpconline.org/digipres/discover-good-practice/tech-watch-reports>).

- The Data Types series of Guidance Notes include titles such as 'Preserving Documents', 'Preserving Spreadsheets' and 'Preserving Email' and provide information on preservation challenges and tips for addressing them.
- The Topical Notes series (linked from the bottom of the web page) are a series of ten publications aimed at a non-specialist audience. They address key digital preservation issues such as storage, file formats and web archiving.



- **The Rapid Assessment Model** (<https://www.dpconline.org/digipres/implement-digipres/dpc-ram>). If you are keen to further your capacity for preserving digital content in house, the Rapid Assessment Model will help you to map where you are and set targets for moving forward.
- **Digital Preservation Procurement Toolkit** (<https://www.dpconline.org/digipres/implement-digipres/procurement-toolkit>). If you are considering procuring a technical infrastructure for digital preservation, this toolkit provides a list of requirements and other helpful tips.

Spotlight on websites

If your business has its own website, this will be a valuable digital record that demonstrates the development and growth of your business and services over time. Websites are typically updated frequently, so preserving them is not a one-off exercise. Snapshots of your website should be captured and preserved periodically.

Website preservation is a complicated technical challenge, particularly for more dynamic websites (for example, where customers can search an online catalogue), but fortunately, there are website preservation experts who can help. Here are some steps you can take to ensure your website is preserved:



- ✓ Find out if your website is already being archived:
 - Check the Internet Archive – simply put your web address in the search box to see if it is already being captured, see <https://web.archive.org/>.
 - Find out whether your National Library or Archive carries out web archiving and whether your website is already being captured.
- ✓ If your website isn't already being archived by these services, consider nominating it for archive.
 - The Internet Archive, based in the United States, has a mission to preserve all websites that are publicly available online. There is a quick and simple way to archive a web page. Add your web address into the box under 'Save page now' and click 'Save' (note that this will save only the exact page that you entered, not your entire website), see <https://web.archive.org/>.
 - Find out whether your website is in scope for preservation by your regional or national archive or library service. Collecting scope and procedures for nominating or submitting a website for archive will vary within different institutions.
- ✓ Be proactive! If you are about to carry out a major update or overhaul of your website, ensure it is archived in its current state before you make any changes.

Get started!

Despite the apparent complexity of digital preservation, many of the steps described in this guide can be implemented quite easily. Consciously attempting to preserve your files for the future will always be better than leaving things to chance, so do start thinking about what digital content you should be looking after for the future and work out the actions you can take now.



Acknowledgements

This guide was originally commissioned by ARCW, funded by The National Archives (UK) through the Covid-19 Archives Fund and created by the Digital Preservation Coalition. It was initially published in English and Welsh by ARCW under an Open Government Licence in August 2022. This revised version of the guide was updated by the DPC in October 2022 for publication to a wider and more international audience and minor updates were made in January 2023. It is published under an Open Government Licence.

The guide draws on existing content originally published as follows:

- The 'Novice to Know-How' learning pathway developed by the Digital Preservation Coalition and funded by The National Archives (UK) as part of the 'Plugged In, Powered Up' Programme.
- The *Digital Preservation Topical Notes Series*, developed by the Digital Preservation Coalition with the support of the National Archives of Ireland.

The illustrations in this guide were created by Jørgen Stamp and are copyright of digitalbevaring.dk and shared under a CC BY 2.5 Denmark licence:

https://creativecommons.org/licenses/by/2.5/dk/deed.en_GB

