Guide to Digital Archiving

by David Parker ©   Version 2 – 11/02/2014

A guide to churches and other small organisation to establish and maintain an archival system for electronic records (including those born digital and those created by digitising existing paper records.

Archiving of digital documents has the same aim as traditional paper archiving - to preserve, manage, and provide access to the selected documentary records of an organisation for as long as they are needed, usually in perpetuity. However, while the aim and general principles are the same, the procedures for digital archiving are rather different, and, in some respects, much more difficult to handle.

Note that one of the difficulties of digital archiving is the sheer volume of digital records produced by an organisation. However, not all items produced are considered important to the official operation of the organisation and therefore do not need to be retained permanently as archival records. (There is some confusion over the terms ‘record’ and ‘document’ - sometimes one or other of these terms may used for all official records or be restricted to those important for retention.)

Digital records provide many advantages over paper records (such as ease, speed and efficiency of searching for information, convenience of distribution, and economy of space), so it is worth persevering despite the difficulties. Furthermore, it is rather regressive to create documents digitally, as is now the universal practice, and then to turn them into paper form for archiving merely to avoid the difficulties of the digital process. There may, however, be good reasons for producing good quality hard copy print-outs of an organisation's most critical documents.

With modest equipment and a few simple policies and procedures, it is possible for a church or other small organisation to create and maintain an effective Digital Archives. Higher level systems will, of course, provide much more functionality and effectiveness. This Guide is aimed at the modest end of the spectrum. For an advanced approach, see your IT consultant, who should be able to recommend and manage a top level enterprise Electronic Document Management System (EDMS) suitable for your needs.

It should be emphasised that 'Digital Archiving' is not the same as backing up or 'archiving' of files in a 'computer' sense. The function of the latter is to be able to restore data in case of hardware breakdown, accidental deletion, corruption or loss of storage media because of fire, flood or theft etc. 'Digital Archiving' refers to the preservation of important digital files for perpetuity and involves management of the files and the provision of ready access to particular files over lengthy periods of time. Note that some EDMS applications are more related to the day by day operation of an office while others may be aimed at archival use where long term retention and accessibility are important.

This document covers the creation and archiving of the main documents of your church such as minutes, reports, financial statements, newsletters, membership rolls, personnel records, correspondence, databases, emails, audio recordings and images (still and motion) etc. (See our other documents on archiving of emails, digitisation of existing paper records and managing church Archives.)

There are five steps in effective Digital Archiving, especially when starting from the beginning.

Step 1 - Audit

The first step is to carry out a full audit of the document creation processes and policies that your organisation is currently employing, and to regularise them where problems are revealed. This step will include consideration of such matters as: Who creates (and revises/updates) documents? Under what authority is this carried out? What procedures are there for regular back-up of documents? What policies govern whether documents are selected for retention by archiving or for disposal? On what computers are they created? On what media are they stored and backed-up? What are the policies and practices for naming of files and folders? What is the folder structure?

What types of file format are used (eg PDF, DOC, TXT, RTF, XLS, JPG, PST etc)? What policies are in place for upgrading hardware and software?

Step 2 – File creation and storage

It is necessary to implement a standardised practice of digital document creation as a regular procedure throughout the organisation, including any needed staff training, hardware and software upgrades, and top level policy determinations. Carrying out the audit of Step 1 is the first stage. Then comes the development of procedures for creation and storage of files.

The aim is to organise the file creation process from the time of initial creation of all documents so that all the vital files that need to be retained over a lengthy period of time (for legal, planning, financial, historical purposes) are created, named and stored in an orderly fashion in safe places on the computer system. It is preferable to create and store vital files in formats that are likely to be accessible for many years to come (which means avoiding highly special proprietary formats in favour of commonly readable ones, or even open source formats that are not dependent on particular software and operating systems.)

All staff involved in creating documents (such as correspondence, reports, rolls, minutes etc) need to be part of this process. There needs to be adequate training and supervision of all staff.

Step 3 - Archiving - the transfer of standardised records to a computerised archiving system

Once a process is established for the creating and storing of files in a systematic manner, the next stage is the archiving process. To understand this step, it will help to make comparisons and contrasts between traditional and Digital archiving

Traditional records, usually kept in well ordered filing systems and substantial Minute and other record books, need to be transferred from the working environment (eg - Minutes of church meetings made by church secretary/administrators and kept in an official Minute book in the Church office) to the archival environment (eg - a stationery cabinet in a suitable room somewhere in the church or at the Baptist Archives) at a stipulated point in time (eg - after 5 years). When this transfer is made, their status changes from current records to archival documents, and they come under the control of the church's Archives.

Now in regard to *digital records* (eg - files on the Administrator's computer or the pastor's laptop/tablet), the same principle applies. In this case, 'transferral' means making the documents which were created by the originating writer accessible and under the control of the Archivist. This typically involves moving the files from one storage medium or area to another. The principle is similar to traditional Archiving, but the process of transferral is not so clearly delineated.

In the traditional system there is one copy of the physical record (eg - the Minute Book) which is moved from one location to another when archived; it can only be accessed in this new location. However, digital records are saved initially to a storage medium 'somewhere' in the computer system; immediately, they can be copied any number of times and sent to many locations (eg - emailed to members of the church, put on a thumb drive for portability, placed on the web site etc). Furthermore, even if there is only one copy of the file, it may be accessible instantly, simultaneously and continuously to many users, depending on the networking facilities of the organisation and the policies implemented by the organisation about access to such documents.

So it is clear that one of the big difficulties of digital record keeping is managing the digital documents - who has access to them? Is there any time constraint? Are there any restrictions imposed on reading, printing, copying, altering, and/or deleting the file, and if so, what policies govern these operations and how are they determined?

On the technical side, there is a potential difficulty in ensuring that the file is what it claims to be (authenticity), and that it has not been altered in any way since it was created/stored initially (integrity). A further issue is whether the original formatting of the document needs to be retained (eg font sizes, headings, margins, colour etc) or whether only the text or graphic content is important. Another problem is obsolescence - whether the file will remain readable over lengthy periods of time (eg - operating systems, software and hardware may change, or the file may become corrupted).

Advanced EDMS software provides sophisticated safeguards for many of these potential difficulties, while good IT support will be needed to overcome others. A lower-level system will need more manual input and operator vigilance to cope with the potential problems.

Automatic or manual transfer

Top level EDMS software is fully automated and integrated with normal office software so that immediately a document is created, it is identified as archival and 'transferred' to the Archives. It is also of course remains available in the normal office environment, with appropriate controls to ensure only authorised staff are able to access it. Parameters to control this transfer process are set up on installation of advanced EDMS software and can be modified from time to time as needed by the System Administrator.

Staff members creating files will need training to ensure that documents are prepared appropriately. Training will also be needed to ensure that the software is operated in such a way that the newly created document is archived properly upon saving as well as being processed and used in the normal manner (eg - sent as a letter, used in a report or becomes part of a publication)

On the other hand, for smaller organisations whose requirements are less, who have modest equipment, and low level IT support, it is possible to set up a budget solution which is almost as effective. This system emulates paper archiving procedures by manually copying or moving the files created in the office environment at a pre-determined time to an archival storage area (such as another hard drive, or an external unit or elsewhere on the network). This process can also be automated to some extent by using software to copy or move selected files to the other area, or else it can be a purely manual operation (Note – this is similar to the familiar practice of backing-up data onto disk or tapes, say, at the end of every week which can be automated by the operating system or carried out manually).

So whether automatically or by manual emulation, the files selected to be archived appear within the Archives environment where they can be preserved, managed and accessed in manner that is functionally similar to the way paper archives are handled.

Top level archival software may, at the time of transfer, include the facility of changing the format of files which are considered to be susceptible to obsolescence to standardised or open formats.

Step 4 - Management of the Digital Archive

At this stage, the files selected for archiving are residing on a dedicated storage medium, having been transferred there according to the official protocols of the organisations (viz - in relation to which files are archived, the time when they are transferred, the formats and folder structures). These archived files now come under the control of the Archives where they are to be managed and provided for access and retrieval under the established policies.

For ensure security, these files should be placed on separate storage area, such as an external drive or on a network drive or some other remote location. These files need to be backed-up in the normal way (ie - regular multiple back-up copies both on-site and off-site). The storage media need to be checked periodically with a view to replacement when necessary, or the refreshment or even migrating the data to a new environment. Regular updating of software and operating systems is advisable. (Note that this is the kind of same procedure that is applied to all electronic data and computer equipment.)

In setting up the protocols for archival storage it is important to store files in formats and locations which are readily accessible on a long-term basis, rather than being restricted to particular software. This means the files should be in formats which can be expected to be accessed well into the future. The files should also be identifiable and accessible individually. So any solution that uses proprietary file formats (which are not likely to be accessible with the passing of time), or one that merges and compresses files into larger units (ZIP, BAK or other similar operations) is not acceptable.

To facilitate searching, retrieval and general management, files should be named and stored in folders which correspond with the original creation of the documents and which are easily identifiable. For example, a file may be named MembersMinutes2005.doc, and it could be stored on your computer in a folder which is named \HopeBaptistChurch\Admin\MembersMeetings\2000-2010. This is better than calling it doc1.doc, and storing it in the default folder, \my documents. Images should be grouped by subject and named individually (eg - rev_jones.jpg in folder \church_opening_2006), although if there are many images of the one event which are somewhat self evident, generic naming of files with a suitably named and dated folder might be sufficient – image0001.jpg, image0002,jpg etc. Advanced systems also provide comprehensive ways to record the description of each document (or meta-data) (such as author, topic, addressee, date of creation etc..

Note that format is important because the layout of a document is as important as the text (ie, fonts, margins, colour etc). Hence it is recommended to preserve the document as a whole rather than just the text. A convenient system to achieve this is PDF/A which embeds fonts and enables the document to read independently of the environment. However, saving of text only may be an acceptable fall back option for some documents. Another option is to print hard copy of critical documents on archival quality paper for normal archiving.

Security of the data from unauthorised use, editing or deletion is another important consideration. It is much easier to read, copy, distribute and print digital material than traditional, and it is also much harder to detect and keep track of these activities. Advanced software has built-in controls to manage these factors, but modest systems need physical and operator controls.

Advanced systems also provide levels of access that can prevent the files being accessed by those without proper authority; this also covers timing so that files can remain inaccessible for stipulated periods of time, except to those with authority otherwise. Manual systems rely on the vigilance and integrity of the Archivist and IT staff.

Access, Searching, Retrieval and Display

Because one of the aims of archiving of records is to provide access to the documents so that they may be readily retrieved at any time in the future, it is important to provide adequate computer facilities and management processes with appropriate overall policies to achieve this.

Data may also be searched for using the normal operating system search functions or similar functions in software. DOC files for example can be searched across a folder or drive by the character string required, restricted by a date (eg - search for the name of person - 'smith' for files created before 2006; or any reference in the whole collection to 'title deeds'). Advanced system provide for searching of meta-data as well as text.

In an advanced system, there are comprehensive ways to retrieve and view data - including display on remote workstations and over intra- or inter-net, or via a mobile device. In a modest system, display possibilities are usually restricted to the original software being used at the local workstation or via a local network.

Policies for access to Digital Archives should follow the same general principles and standards as for traditional archiving - with due allowance for differences in digital environment. Furthermore, while normal security measures should be taken, it is important not to tie up the files so tightly that they become inaccessible and unusable when there are changes to personnel, hardware and software.

Cloud computing

Cloud computing provides advantages for mobility and access. While it is convenient for staff in any location to create and save files initially in the cloud, the organisation needs established policies place which ensure that any and all official files are saved to locations which are accessible to and controlled by the organisation itself and not restricted to personal locations. Care needs to be taken regarding security and access of files stored in the cloud.

Storage of archival data in the cloud may be a good solution for providing an off-site alternative, but it should not be the only storage site because of larger questions about its long-term security and accessibility. If the cloud is adopted as a secondary storage location, care needs to be taken to avoid complications due the presence of multiple copies of data.

Emails

See our Guide to Emails for additional information. The following assumes the use of MS Outlook, but other email clients will provide similar features.

Emails can be difficult to archive, and the number of emails generated by an organisation is also a limiting factor. But the complexities can be reduced with careful planning and operations. First of all, cull emails so that only important ones are retained. Then use 'rules' to place emails into various categories. Also use the 'archive' feature to create separate PST files. These PST files can now be archived according to the procedures set out in this Guide. If necessary, they can be converted into other formats (such as text) for long-term access.

Digitisation of Paper and other records

Files that have been created by digitisation of paper and other records can be included in the scheme of digital archiving described in this Guide. Once created they are treated in the same way as a born-digital record. Thus, all digital records are handled together in integrated archival collection of the organisation's records.

See our Digitisation Guide for more information

Step 5 – Documentation, Training and Review

Once the system of digital archiving has been established it is important to document it fully. This will include details of the overall policies, structure or architecture of the system, the hardware and software involved, methods of operation. This documentation should become part of the organisation's standard procedure manual. Staff should be trained to operate the system, and provision for regular reporting on activities incorporated into the normal life of the organisation. Finally, the system should be reviewed regularly with a view to improvement.

Further reference:

Contact the Baptist Church Archives Qld for more details on this Guide (archives@qb.com.au).

Refer to our other Guides for information (Managing Church Archives; Digitisation; Emails)

http://parker.org.au/d-MangChrchArc.pdf http://parker.org.au/dig-guide-c.pdf http://parker.org.au/d-emails.pdf

Many government, university and corporate organisations provide information in the internet about archival principles, procedures and products. Note that many solutions offered are highly sophisticated, requiring a large budget and strong IT support, and also operate within a complex legislative framework.

State archives in Australia include Queensland State Archives, Public Record Office of Victoria (PROV), State Records in NSW, South Australia and Western Australia. See also the official government archives of UK and USA etc.

PROV offers assessment of commercially available EDMS software at
http://prov.vic.gov.au/government/vers/implementing-vers/vendor-assessment/product-compliance-status
(accessed 18/12/2013)

HOPPLA - Home and Office Painless Persistent Long-term Archiving system – version 2.1 - a free application suitable for church use - visit http://www.ifs.tuwien.ac.at/dp/hoppla/ (Note that this application is no longer being developed, but the current version 2.1 is stable, well featured and easy to use.)