Canberra TMG Users Group

Archiving your data electronically

Prepared by Shane Baker
for the March 2006 meeting.
Revision 1.2 – revised 7 May 2006

Archiving your data electronically

Why Archive?

For the purposes of this session (and paper), *archiving* refers to making copies of your data (including exhibits) which:

- are secure from damage,
- are likely to survive uncorrupted for an extended period,
- can be easily retrieved, and
- are reasonably proofed against changing technology.

Why archive electronically?

Basically, because with care, one can make copies which are robust, and which can be read into your computer with a minimum of effort.

Many people argue that paper is the best medium for archiving. It's certainly true that it's a proven medium. It doesn't deteriorate over time (provided acid-free paper is used and it is stored in appropriate conditions), and it can be retrieved any where, any time – even during a power failure.

However, paper has its limits:

- Not all paper is created equal. Look at a newspaper clipping or even paperback book after a few years.
- Paper is bulky. Print out your entire database and see!
- Paper is organic. Insects, small animals and fungus all can and will eat it – and it burns beautifully.
- It's not a medium for storing multimedia.
- Errors can and will occur if you have to re-enter your data.

Archiving your TMG data

Automatic Archiving

TMG has been written with its own archiving routine which will "nag" the user each time TMG is closed down.

If your copy doesn't have this activated, do it tonight. In TMG, go to File / Preferences and under *Program Options* select *Startup and Exit* (see Figure 1). Ensure *Prompt for backup* has a tick next to it.

Next, be sure you know where your backup is going. For the most secure outcome, select a physically separate hard drive to where your TMG data is being stored. A separate partition is a second best option. To do this, go to File / Preferences and under *Current project options* select *Advanced* (see Figure 2). Use the usual Windows "box with 3 dots" method to select a place for your backups.

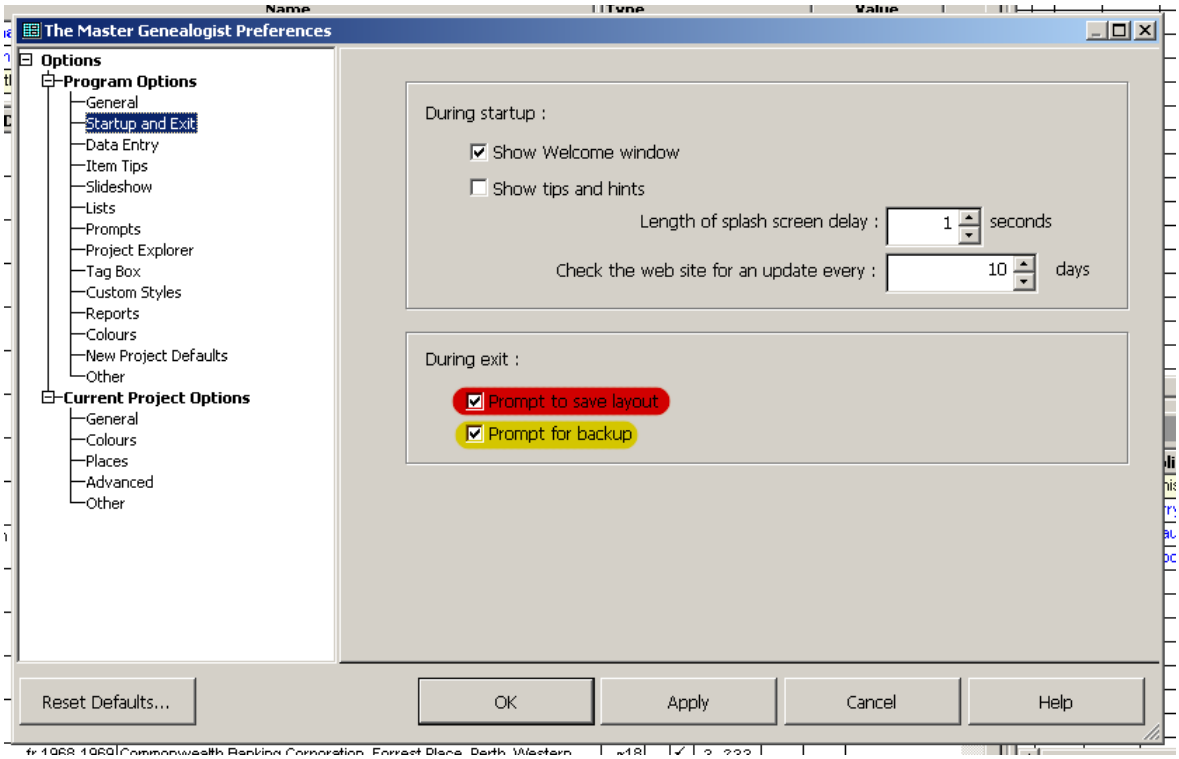

Figure 1

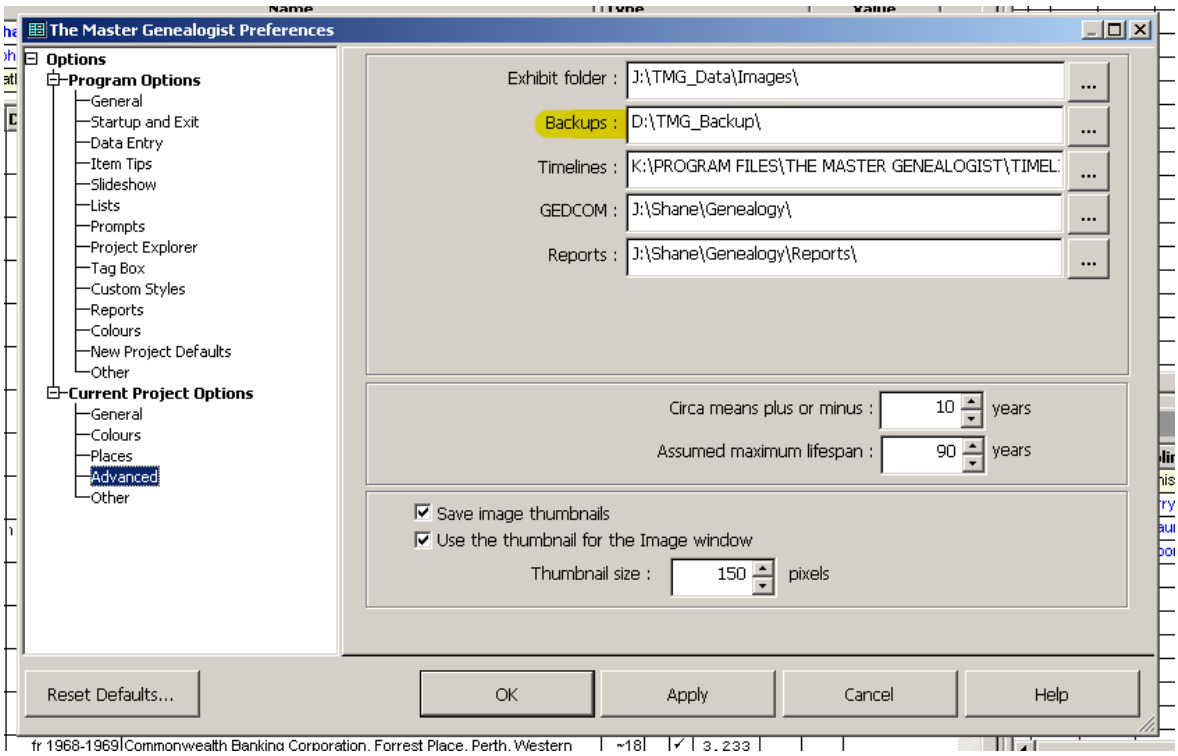

Figure 2

However, before you archive anything, it's worth taking a couple of moments to ensure your data is robust. Go to File/Maintenance (see Figure 3) and at least validate and reindex your data.

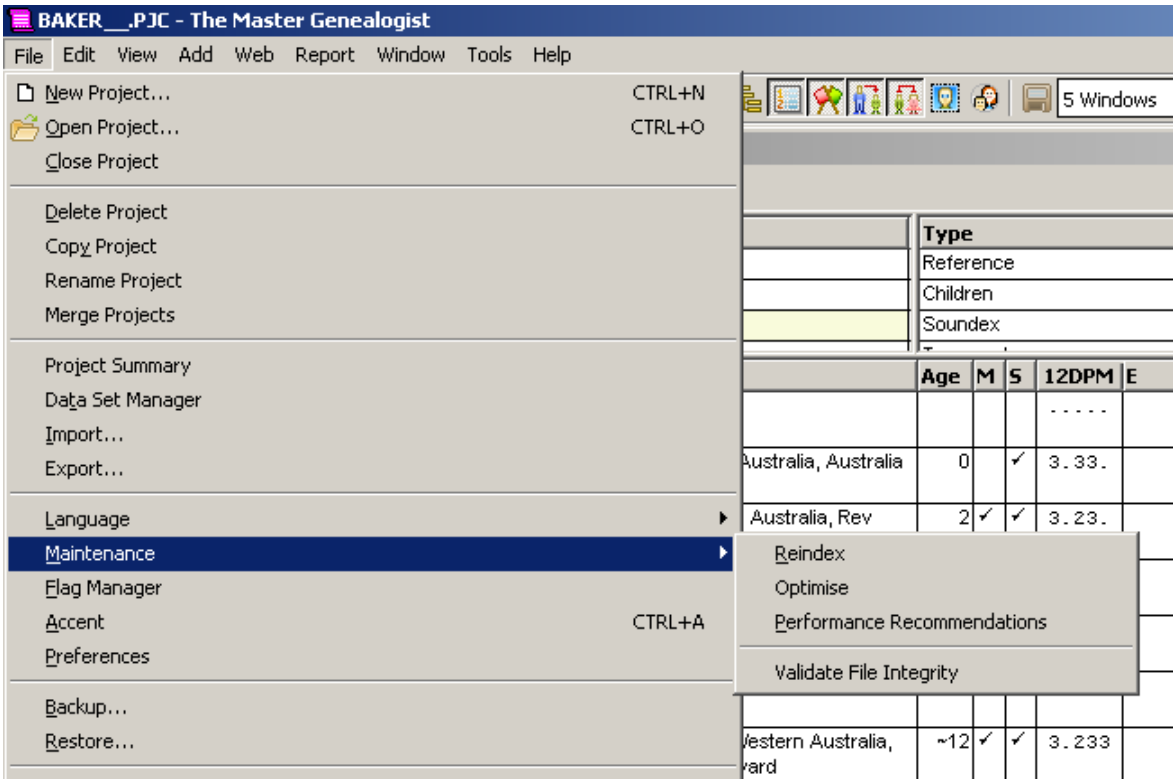

Figure 3

After changing your settings, TMG will prompt you to archive your data each time it is shut down – and your archive file will be stored at the location you have specified.

Archiving in TMG is pretty easy. It comes with appropriate options – or (as with most things in TMG) you can make up your own archive criteria.

Upon starting a backup, you will encounter the *TMG Backup Wizard* (see Figure 4). As you see, several backup configurations are offered, with a text box explaining what it does. Alternately, you can chose to create a new configuration.

After selecting a configuration, clicking *next* takes you to Step 2 (see Figure 5). This shows you what you are backing up and where your backup is being stored. It also gives you the option of naming the backup with a unique name comprising data set name, date and time (perfect for a backup that will be stored off-line), or simply cycling through a series of numbered backups. (In this case, it cycles through 10). If you chose that option, the old reliable "are you sure you really want to do this" box comes up. If you're really sure, click yes.

Step 3 of the process (see Figure 6) provides many options for items to be included in your backup. While including everything is the safest option, it naturally increases the size of your backup file and (for slower machines at least) may increase the time taken for the backup. That said, when backing up is concerned – can you be too conservative?

It should be noted that Chapter 2 of *Getting the most out of The Master Genealogist* warns us that TMG does not preserve a number of files used by TMG to preserve your personal preferences and enhancements. I cannot comment on this, though of course I would tend to defer to Hoffman et al on this as in all things. I suggest you consult that reference if concerned about this issue.

I can confirm that I have found it possible to backup my entire set of TMG program files plus data and exhibits on one CD.

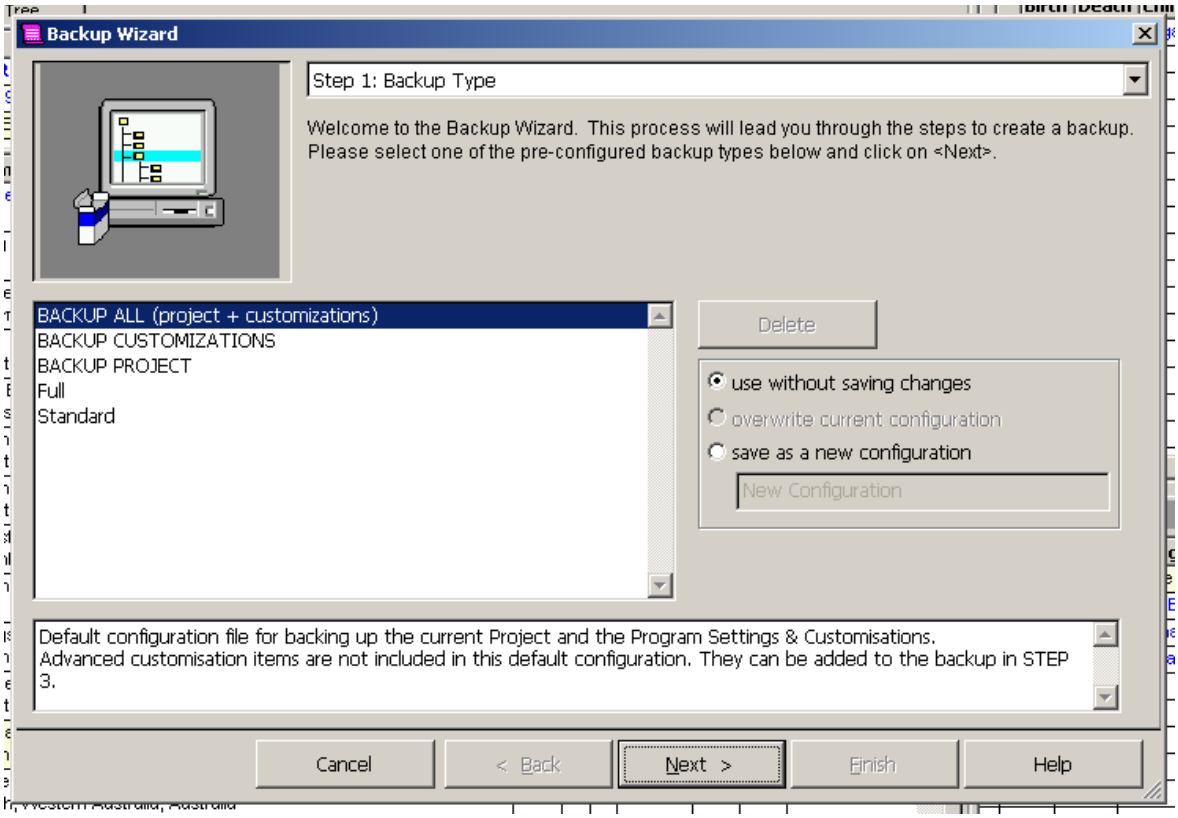

Figure 4

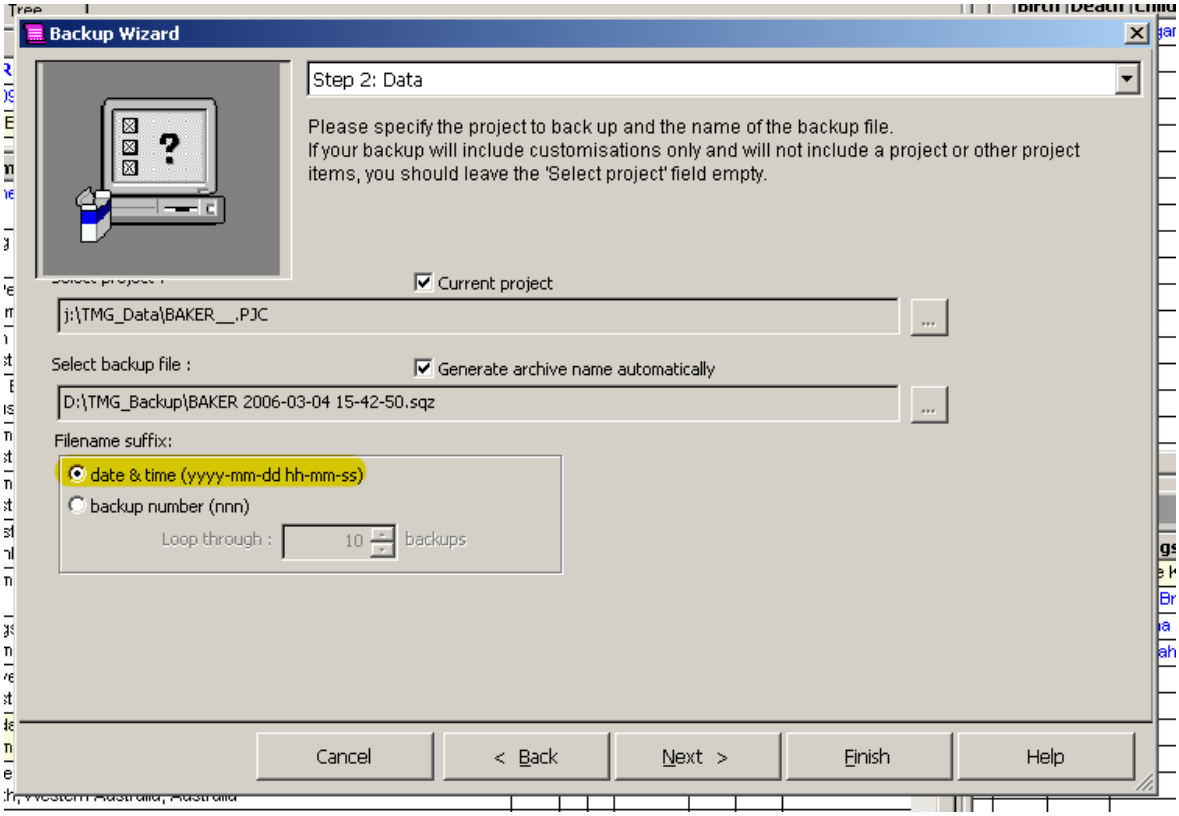

Figure 5

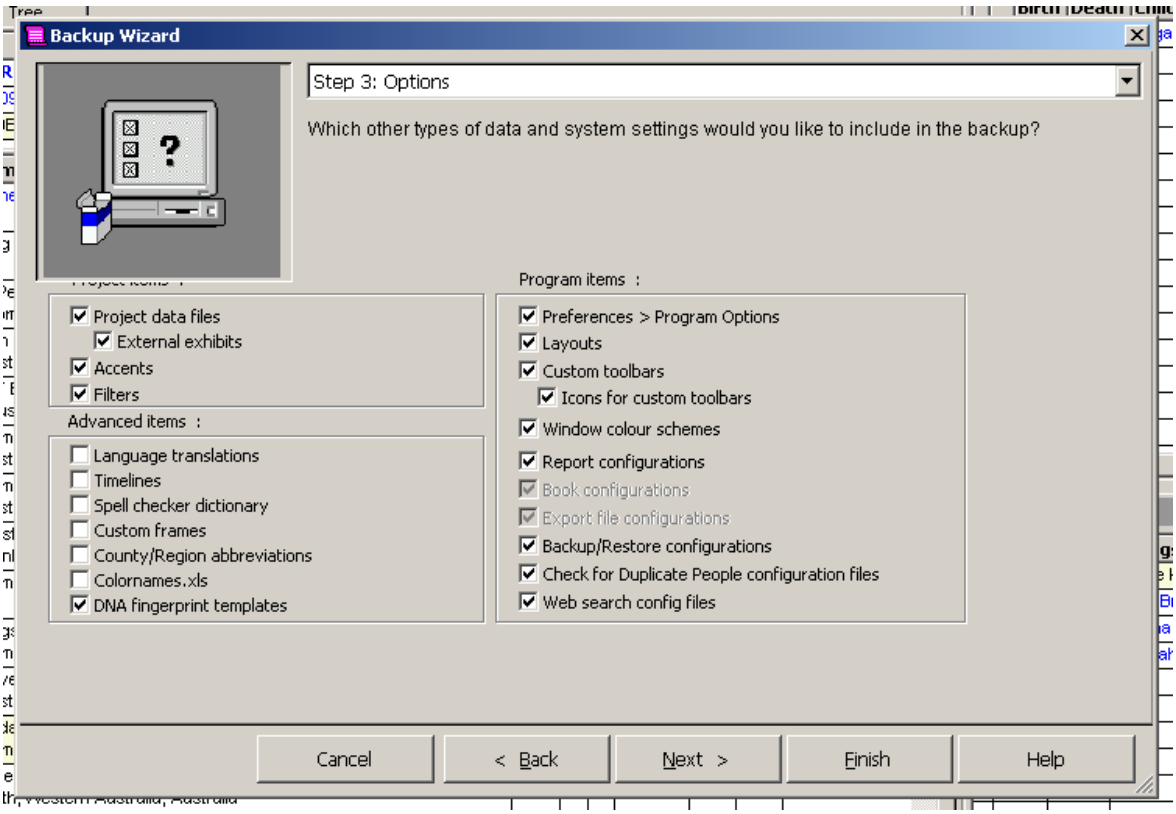

Figure 6

Next (if you've included them to be backed up), you will be asked which layouts you wish to archive (see Figure 7). Logically, you should select *All*.

The following screens offer similar options for *Custom Toolbars* and *Report Configurations*.

Step 7 of the process offers various levels of compression of your data – from none to maximum. Each will offer different advantages and disadvantages: the maximum compression takes the longest – and vis versa. Being a simple soul, I stick with *Normal*.

Step 8 is *Finish*. This gives you the options of cancelling, going back in the wizard, or Finish – which will start the archiving process. In fact, one can click *Finish* on any of the previous screens and TMG will use its default settings.

Archiving Manually

You can also chose to archive at any time. File / Backup will get you into the wizard.

Other options

Of course, data can be exported and archived in a number of other ways. Creating a web site with *Second Site* is one way of preserving data, as is making a complete hard copy of your data via reports.

Other electronic options can be found in the *export* menu found at File / Export. Probably the best option here is to chose GEDCOM – which can not only be imported by TMG, but also by almost any other genealogy program. The downside is that GEDCOM does not store all the data held by TMG, and Wholly Genes "strongly discourages" the use of GEDCOM. That said, GEDCOM is the *lingua franca* of genealogy programs.

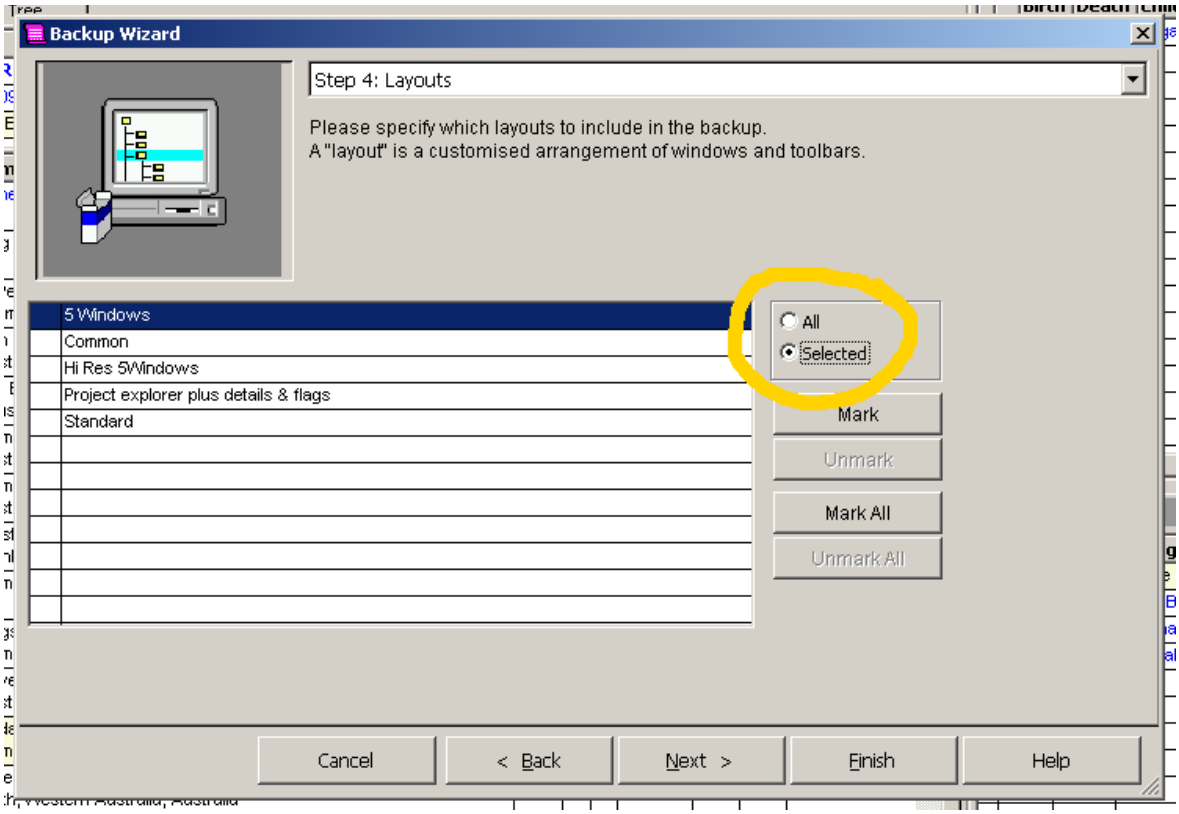

Figure 7

Archiving Exhibits

Although (with the appropriate settings) TMG also archives our exhibits, it may be desirable to archive them separately – or you may simply wish to preserve other materials which are not included in your TMG database.

Text

As anyone who has used computers over a number of years can confirm: word processor file formats are not stable. I personally have a number of documents created with obscure word processors which I can no longer open. Even sticking with one brand of word processor is no guarantee; Microsoft in particular has an appalling track record of unsupported changes to its own file formats.

So, what format is best to store text? The safest file format is **plain txt** (.txt). The downside of this format is that it contains no formatting. Convert a document to text, and you will only have characters, tabs and returns – with no bolding, font choices or type size preserved.

Another option is **RTF** or *Rich Text Format*. Although this format can be read and edited by practically all word processors and free editors (such as Microsoft's *WordPad*), and it carries most of the formatting one needs to present text in an acceptable way, it is a proprietary format (i.e.: owned by Microsoft). Hence, while it is probably the best option for the foreseeable future, whether it is an archival format is questionable.

Possibly the ideal option is a word processor which stores data in the ***XML*** format. Why XML? Well, it was developed by the same team which developed the familiar HTML used on the World Wide Web.

According to the National Archives[1]:

> The National Archives' approach to the preservation of digital records involves converting records in proprietary data formats to equivalent data formats in XML.

The European Union has also adopted XML as its preferred format, and consultants Forrester Research[2] has predicted a move to XML for office documents. Indeed, in May 2006, the International Standards Organisation announced that the format is *the* standard for word processed documents under its standard ISO 26300.

A *free* XML-compliant office suite called ***Open Office*** is available[3] from OpenOffice.org .

To summarise: text is the safest very long-term option, but XML appears to be the best option for storing data with formatting.

Images

Increasingly, people are seeing image files as part of their family heritage, and are wishing to preserve them. The 2003 Canberra fires showed how easily a collection of irreplaceable family photos can be lost, while the move to colour photos in the 1960s and 70s is resulting in photo albums of brown, barely recognisable images.

One of the best options for addressing this is to scan images and store them digitally. (This could include key documents, as well as photos, etc.) While it is true that an image file isn't the same as an original wedding photograph from the early 1900s, it has the advantage of being reproducible and easily shared with family in other locations.

File Format Choices

Once an image has been digitised, it must be saved in an appropriate format. There is a bewildering range of options, but in practice, only a few are likely to be used by you unless you have a special need.

As with most things in life, in image files, there's no such thing as a free lunch. An image file format that is relatively small is that size because it's disposed of some data. The following table provides a size comparison with several different image file formats. They are the same image, displaying the same size. The difference is the file format:

1 http://www.naa.gov.au/recordkeeping/preservation/digital/xml_data_formats.html
2 http://www.forrester.com/Research/Document/Excerpt/0,7211,36355,00.html
3 http://download.openoffice.org

Image File Type	File Size (k bytes)
PhotoShop (.PSD) (Adobe proprietary)	17,647
TIFF	17,647
BMP (Windows proprietary)	17,626
TIFF (compressed)	10,966
RAW (Nikon)	6,049
JPEG (maximum quality)	4,371
JPEG (medium quality)	799

TIFF

TIFF is a widely accepted, non-proprietary format, and the one you should use if you ever send the file to a commercial printer.

TIFF is a non "lossy" format – and data is not lost. Files, however can be large.

TIFF is the only real option for storing digital images for archive purposes because it is non-lossy and non-proprietary. Compressed TIFF will save space, but may not be as compatible as a non-compressed file.

JPEG

JPEG is the format in which most (but not all) digital cameras download their files to a computer. The files they produce are relatively small, with a good colour range. (For this reason, JPEGs find a lot of use on the internet.)

However, JPEGs keep themselves small by discarding data. As such, unless you have a good reason to do so, or are preparing files for a web page, don't use JPEGs. If you must use JPEG, save it at the maximum quality setting.

GIF

GIF files also find much use on the internet, again because of their small size. Their main disadvantage is that they store very little colour data, and are therefore used mainly for the icons and buttons seen on some web sites. *Never* use this format for photos.

"Raw" Files

Most up-market cameras offer RAW as a means to download images to your camera. Raw is not so much a file type as a simple "dump" of pixels from the camera's image sensor. As such, it can be problematic to use a RAW image without either proprietary software. Adobe is offering[4] their Digital Negative (DNG) format as an agreed common format for camera image files. While there are sound reasons why this initiative should be successful, it is too early to place all your eggs in that basket.

While Raw will probably give you the best results from your camera, save as TIFF (and possibly DNG as well) before editing or archiving the image.

4 http://www.adobe.com/products/dng/index.html

Others

Other common formats are

- "Ping" files (.png) are finding increased use on the internet. Again, a lossy format,
- Windows Bitmap files (.bmp) which will be accepted by most Windows applications, but which are proprietary, and
- the proprietary Photoshop format (.psd).

Archiving Media

How long your archives last will rely on three factors:

- your choice of format;
- your choice of medium; and
- the storage conditions.

Paper

Data

Acid free paper stored correctly (see below) will last and last. Making hard copies of your data will preserve it against technical change indefinitely. However, data on paper is fragile, bulky and not easily copied.

Newsprint is, as we all know, subject to browning within weeks, This is due to its high acid content. If a newspaper clipping is worth keeping, it's worth copying to another format.

Images

As anyone will know from comparing a 1900 black and white (silver halide) photo with a 1975 colour photo, a properly printed and processed black and white image will last for hundreds of years. A colour image will begin to deteriorate after only a few years.

Recent colour emulsions, especially from *Fuji*, have better archival qualities. However, colour isn't really an option for long-term option.

If "hard" colour copies are really necessary, probably Kodak's *Kodachrome* slide film, and Epson's *Ultrachrome* printer inks and papers are your best option. (Some brands of ink and paper have appallingly short lives. It would be well to check their archival qualities before relying on many digital printing systems.)

Hard Discs

Having digitised your data by keying it into TMG or a word processor, scanning it, or taking digital photos, how do your keep your files safe?

Initially, we all store them on our "trusty" hard disk drive (HDD). HDDs are cheap (in March 2006, I bought a 250 gigabyte drive for $A187 – delivered), fast and generally reliable. However, they are not infallible.

Modern external drives (connecting via USB or Firewire) are a useful secondary backup and can be used to transfer large amounts of data between computers. However, inside that case is still a HDD – with the added complication that it can be dropped, lost or stolen.

See your hard disk as an interim storage solution at best.

CDs and DVDs

When CDs were first launched, they were presented as archival in nature and almost indestructible. They aren't – on either count. Apart from the risk of scratching, the choice of reflective material (gold, silver or other alloys) and choice of dyes impacts heavily on potential life.

Furthermore, as with all electronic technology, they have a finite technical life. Even as DVDs are replacing CDs for data storage, the next generation of DVDs is about to be launched commercially. However, CDs and DVDs have several advantages over other media. They are:

- cheap;
- small;
- easy to copy – either for sharing with others (thus improving chances of survival of some natural disaster) or to be copied to the next generation of digital media; and
- because the data they store is digital, the information doesn't degrade when copied. (We have all seen seen the degradation of a photocopy of a photocopy. This is because it is an *analogue* copy. By comparison, a file containing a scanned (digital) version of that document can be copied over and over again.)

However, like oils, media ain't media – you get what you pay for. Those cheapie CDs or DVDs in the local supermarket or computer fair may last a year or less. Even among known brands, some have a better reputation than others. Notably, Taiyo Yuden and ProDisc (Mitsui) have proven track records. ProDisc offers gold CDs claimed to be good for 100 – 300 years – at $3.25 each, while their gold DVDs ($4.95) are claimed to be good for up to 100 years. Second options are the brands you recognise: Imation, TDK, Verbatim, Fuji, etc.

Having bought quality disks, there are a number of important things to keep in mind:

- use good burning software. Products like Nero (shipped with many CD and DVD burners) work well;
- Burn the disc as slowly as the software will allow. This seems to produce a better, more stable disc;
- Never use adhesive labels or harsh permanent markers on your disc. Print (on printable discs) using a high quality printer or use markers made for the purpose from TDK etc.
- Store the disks like vinyl – vertically, away from light and heat, and in high quality cases. Companies such as ProDisc sell archive quality cases for around 68¢ each.
- Make copies – and store them in separate locations.
- Check the discs from time to time. Programs like *CD Check* will check the data on both CDs and DVDs.
- If possible, use CDs in preference to DVDs. For technical reasons, CDs are more robust than DVDs. But a high quality DVD may still out-perform a cheap and nasty CD!
- Transfer data to new forms of media before technological obsolescence makes copying impossible. (Remember 5¼ inch floppy disks?)

Points about storing your archives

Pretty much any record, regardless of medium will last longer if you avoid:

- moisture;
- heat;
- wide temperature variations;
- light;
- fire;
- insects; rodents and other vermin;
- chemicals – especially acidic contaminants; and
- reliance on a single copy – in a single location.

Remember: Archive often!!

Useful References

TMG

Hoffman, Lee H. (ed) *Getting the most out of The Master Genealogist (Australian edition)*. Gould Genealogy, Modbury North, SA. 2003

Text Editing

National Archives of Australia - *Recordkeeping - Digital Preservation - XML Data Formats* - http://www.naa.gov.au/recordkeeping/preservation/digital/xml_data_formats.html

Forrester Research *XML And PDF-A Will Reign As Future Document Archive Formats* March 2005 - http://www.forrester.com/Research/Document/Excerpt/0,7211,36355,00.html

Image Storage

Australian Photography magazine, *Storage and Archiving*, February 2006

Wilhelm Imaging Research, Inc. - http://www.wilhelm-research.com/

CD/DVD Media

Australian Personal Computer magazine: *Stop the rot* - http://www.prodisc.com.au/files/apc_article.pdf

Wall Street Journal: *Beware the fading dye* - http://www.prodisc.com.au/files/wsj_article.pdf

Sources

CD/DVD Media[§]

ProDisc: http://www.prodisc.com.au

Taiyo Yuden: PCX Pty Ltd - http://www.pcx.com.au/

Software

CD Check - http://www.elpros.si/CDCheck/news.php

Open Office - http://www.openoffice.org/about_us/introduction.html

Archival Materials[§]

Conservation Resources – http://www.conservationresources.com.au

Zetta Florence - http://www.zettaflorence.com

§ NB: These companies may not be the only source of these materials. Furthermore, this is not an endorsement of the companies, and you are urged to investigate alternate products and sources.

