


Saving Data Online for Future Generations

Hope N. Tillman & Walt Howe
Essex Society of Genealogists
April 16, 2011

- ▶ We want the genealogical information we have created to survive after us so that it is findable by those who will be interested.
 - ▶ Your choice of what and how you put data online and where you put it online should meet your purpose:
 - Accessibility
 - Permanence
 - Correction handling
- 

You need a Personal Preservation Plan!



158	15	Essie Ordway Taylor	F	S	6	3	28	Diphtheria	
159	23	Frances A. Hamlin	Wid	M	39	-	-	Consumption	
160	26	Sarah Edith Taylor	F	S	-	4	-	Indigestion	
161	30	James Garney	M	S	-	7	-	Cholera Infantum	
162	30	Joseph Perford	M	S	-	7	-	Diphtheria	
163	Oct 24	Jacet Alfred Chinn	M	S	15	7	17	Diphtheria	
164	5	Mary Ann Graham	F	S	25	-	-	Consumption	
165	6	Mary Gray	F	S	-	2	-	Diphtheria	
166	10	John Richmond	M	M	77	-	-	Old Age	
167	16	Mary Hayman	F	S	18	5	14	Consumption	
168	16	Wesley Nelson	F	M	54	10	-	Diphtheria	
169	18	Ellen O'Brien	F	M	24	9	-	Diphtheria	
170	19	Jane Gray	F	S	18	2	11	Cold	

Personal Preservation Plan

1. Make a plan
 - Ask yourself what you want to save and why?
2. Associate metadata with your data
 - Metadata is information describing data
 - **An image** may include metadata that describes how large the picture is, the color depth, the image resolution, when the image was created, and other data.
 - **A text document's** metadata may contain information about how long the document is, who the author is, when the document was written, where the document is located, and a short summary of the document.
 - Tag documents, images, etc. appropriately and consistently
 - By date
 - By surname
 - By relationship
 - By keyword

Personal Preservation Plan

1. Make a plan.
2. Associate metadata with your data.
3. Identify the data you need to backup
 1. Data from your genealogy software
 2. Scanned Images
 3. Scanned Documents (including scrapbooks, storytelling)
 4. Emails
 5. Audio/video files
 6. Research logs
 7. Internet bookmarks
 8. What else?

Personal Preservation Plan

1. Make a plan.
2. Associate metadata with your data.
3. Identify the data you need to backup
4. Use multiple approaches and make multiple copies –
Follow the LOCKSS model (<http://lockss.stanford.edu>) Lots of Copies Keep Stuff Safe
 - **For now:** Backup files from your computer to both (1) **External hard drive** and (2) **Cloud backup(s)**. (3) Use **Flash drives** or **DVD/CDs** for current project work
 - **Long term:** Ancestry, FamilySearch

LOCKSS

- ▶ Backup to flash drives
- ▶ Backup to external HDD
- ▶ Back up to cloud service

External Hard Drives

- ▶ CNET Reviews: <http://reviews.cnet.com/best-external-storage-drives/>

The screenshot shows a CNET website page with a dark header. The main navigation bar includes 'Reviews', 'News', 'Downloads', and 'Video'. Below this is a secondary navigation bar with categories like 'Cell Phones', 'Camcorders', 'Digital Cameras', 'Laptops', 'TVs', 'Car Tech', 'All Categories', and 'Forums'. The article title is 'Best external storage drives and hard drives', updated on May 25, 2010. The author is Dong Ngo, a Senior Associate Editor. The article text describes the purposes of external hard drives and lists connection types. A featured product is the Clickfree C2N (500GB), priced at \$98.00 - \$117.04, with a 4.5-star rating and 'Excellent' review. Another product, the Seagate FreeAgent GoFlex Pro (500GB), is partially visible at the bottom. The footer contains social media links and site navigation options.

Home > Reviews > Best 5 storage devices > Best external storage drives and hard drives

Best external storage drives and hard drives

(Updated May 25, 2010 7:09 AM PDT)

An external hard drive has three main purposes: to expand your computer's storage capacity, to backup your data and to share data between computers. It's also easy to use. Most of the time, you just need to plug it into the computer and use it as an internal hard drive. There are desktop external hard drives based on the 3.5-inch internal hard drives and laptop (or portable) external hard drives that are based on the 2.5-inch internal hard drives. Generally, external hard drives are connected to a computer using collectively these types of connections: USB 3.0, USB 2.0, FireWire 400, FireWire 800, and eSATA. Portable external hard drives are also often bus-powered, meaning they require only one cable for both data and power connections. Here's the list of our current favorites. This list is updated regularly as more devices are reviewed.

Clickfree C2N (500GB)
The Clickfree C2N makes an excellent backup solution for home users, including those with multiple network computers.
Price: **\$98.00 - \$117.04** (check prices)
★★★★★
Excellent
[Read full review >](#)

Seagate FreeAgent GoFlex Pro (500GB)
Though other external hard drives offer a preset menu of ports, the Seagate FreeAgent GoFlex cuts

Cloud Backup

See Wikipedia table comparing players on such features as platform, encryption, type of plan, max file size and whether it supports features such as encryption, syncing, auto-detecting changes:
http://en.wikipedia.org/wiki/Comparison_of_online_backup_services

Current major players include:

- ▶ Backupify: <http://www.backupify.com>
- ▶ Carbonite: <http://www.carbonite.com>
- ▶ Dropbox: <http://www.dropbox.com>
- ▶ Idrive: <http://www.idrive.com>
- ▶ Mozy: <http://www.mozy.com>
- ▶ Sugarsync: <http://www.sugarsync.com>
- ▶ Syncplicity: <http://www.syncplicity.com>
- ▶ Windows Live Skydrive: <http://skydrive.live.com>

Market shakeout continues!

Linking to External Documents versus Creating Copies of External Documents

- ▶ **Consider whether to link to external documents or to create copies of external documents or whether you should do both:**
 - How permanent are your links?
 - How current are your copies?
 - What versions do you have?
 - Do you have the best/newest version of the data?
 - i.e. improved census images

Personal Preservation Plan

1. Make a plan.
2. Associate metadata with your data.
3. Identify the data you need to backup
4. Use multiple approaches and make multiple copies
5. **Plan for addition of data elements to save**
 - Consider recent additions: email, web address, DNA
 - What will be next?

Personal Preservation Plan

1. Make a plan.
2. Associate metadata with your data.
3. Identify the data you need to backup
4. Use multiple approaches and make multiple copies
5. Plan for addition of data elements to save
6. **Plan for the transition of digital data to new storage media.**
 - Migrate before technology goes away
 - Migration must be an ongoing process

Personal Preservation Plan

1. Make a plan.
2. Associate metadata with your data.
3. Identify the data you need to backup
4. Use multiple approaches and make multiple copies
5. Plan for addition of data elements to save
6. Plan for the transition of digital data to new storage media.
7. **Plan for transitions in data stewardship: pick trustworthy repositories.**
 - Role for interested family members.

Personal Preservation Plan

1. Make a plan.
2. Associate metadata with your data.
3. Identify the data you need to backup
4. Use multiple approaches and make multiple copies
5. Plan for addition of data elements to save
6. Plan for the transition of digital data to new storage media.
7. Plan for transitions in data stewardship: pick trustworthy repositories.
8. **Take into consideration privacy and security**
 - Are you protecting the privacy of living people?
 - Are you sharing all that you can?

Personal Preservation Plan

1. Make a plan.
2. Associate metadata with your data.
3. Identify the data you need to backup
4. Use multiple approaches and make multiple copies
5. Plan for addition of data elements to save
6. Plan for the transition of digital data to new storage media.
7. Plan for transitions in data stewardship: pick trustworthy repositories.
8. Take into consideration privacy and security
9. **Review and update plan regularly**
 - Review goals
 - Review content and privacy
 - Review technology

Your choice of genealogy software

- ▶ [Family Tree Maker](#) – PC and MAC
- ▶ [Personal Ancestral File \(PAF\)](#) – PC *
- ▶ [The Master Genealogist \(TMG\)](#) – PC *
- ▶ [Legacy](#) – PC *
- ▶ [Roots Magic](#) – PC *
- ▶ [Ancestral Quest](#) – PC *
- ▶ [Family Historian](#) – PC *
- ▶ [Reunion](#) – MAC
- ▶ [MacFamilyTree](#) - MAC
- ▶ Others. *This is just a sampling of the major packages.*

* on MAC with Windows emulator

Genealogy Software Choices

Your choice of genealogy software

- ▶ Family Tree Maker
 - ▶ Personal Ancestral File (PAF)
 - ▶ The Master Genealogist (TMG)
 - ▶ Legacy
 - ▶ Roots Magic
 - ▶ Ancestral Quest
 - ▶ Family Historian
 - ▶ Reunion
 - ▶ MacFamilyTree
 - ▶ others
- ▶ Choose a package with a comfortable interface for you
 - ▶ Decide what kind of sharing is right for you?
 - ▶ Update your software regularly
 - ▶ Be prepared to change your software, as companies are acquired or leave the marketplace. Market shakeout continues.

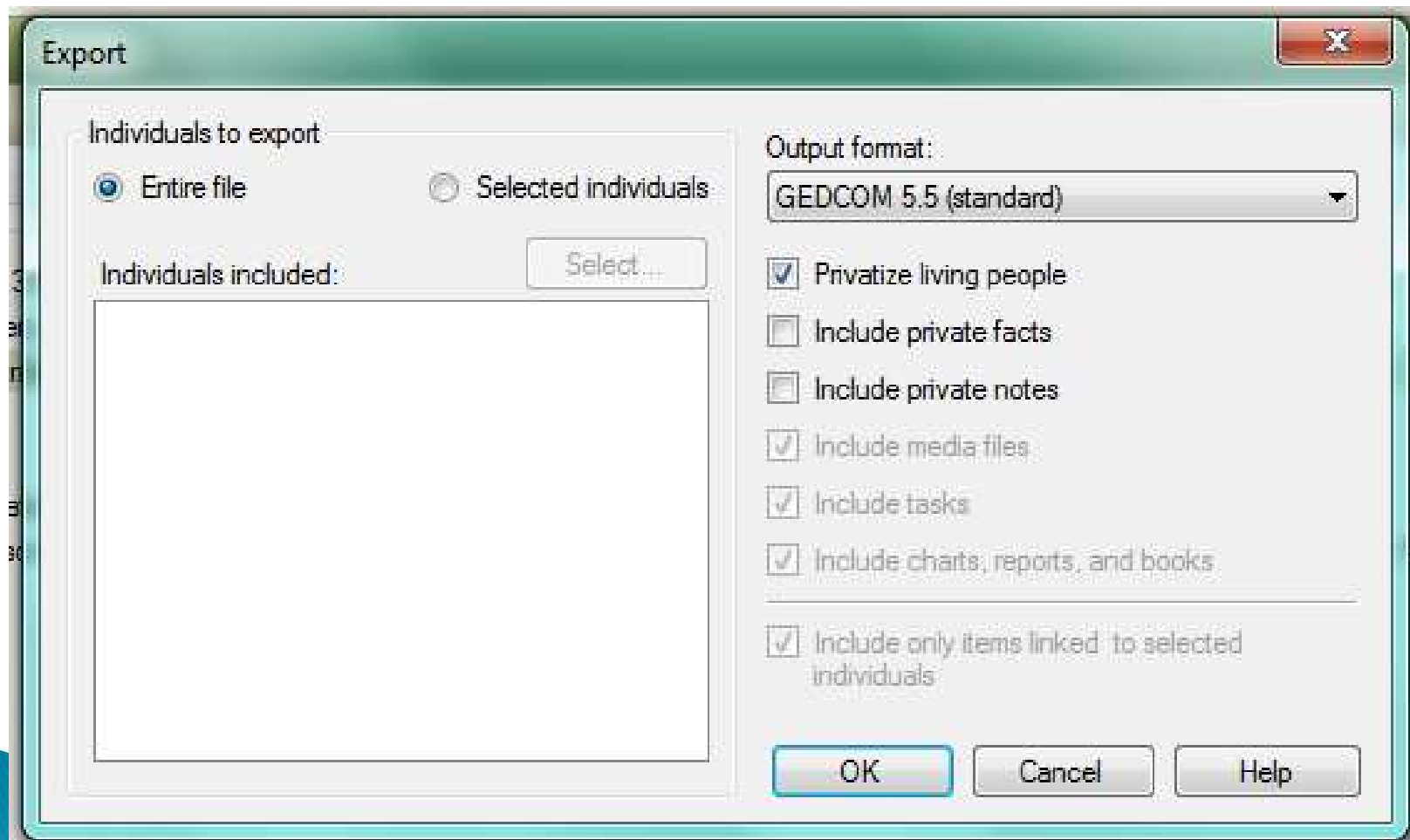
Genealogy Software Choices

Questions to Consider

Your choice of genealogy software

- ▶ Pick a product that supports standards for
 - **Display (character sets)** – , i.e. non-English alphabet needs
 - **Output/input (gedcom, xml)** – Most major packages have gedcom capabilities. More about this later
 - **Sources/citations** – Following standard practices to identify the sources you have used is critical to your work.
- ▶ Decide whether your software will be used solely to **collect and record your data** or will also **serve as a research tool**, with access to databases within it.

All major software packages support import and export of GEDCOMs. You can create a GEDCOM file for your records by exporting from your genealogy software.



GEDCOM

Definitions:

- ▶ GEDCOM, an acronym for GENEalogical Data COMmunication, is a de facto specification for exchanging genealogical data between different genealogy software. GEDCOM was developed by The Church of Jesus Christ of Latter-day Saints in 1984 as an aid to genealogical research.
<http://en.wikipedia.org/wiki/GEDCOM/>
- ▶ A type of file format used by genealogical software to share a family tree.
Some DNA testing companies, databases, and projects allow a user to upload and attach a GEDCOM file to results, such as Ysearch and mitosearch.
http://www.rootsweb.ancestry.com/~genetic_genealogy/glossary.html
- ▶ Various versions of GEDCOM: Current version most heavily used is Version 5.5 (supports Unicode character set – not in earlier versions)

GEDCOM Tags

- ▶ GEDCOM uses a lineage-linked data model based on the nuclear family and the individual. A GEDCOM file consists of
 - **Header Section** (HEAD) (describes the source of the program)
 - **Submitter Section** (SUBM) (identifies the submitter)
 - **Records**: people (INDI records), families (FAM records), sources (SOUR records), other miscellaneous records including notes, repositories, objects (NOTE, REPO, OBJE)
 - Individual records (INDI) define individuals – may be many lines long
 - The family record (FAM) links the HUSB, WIFE, CHIL by their ID numbers.
 - **Trailer Section** (TRLR)
- ▶ Every line of a GEDCOM file begins with a level number
 - Top-level records (HEAD, TRLR, and each INDI, FAM, OBJE, NOTE, REPO, SOUR, SUBM) begin with a line with level 0, while other level numbers are positive integers.
- ▶ GEDCOM allows multiple names for a person, multiple events (e.g., 2 birth dates with one preferred and both sources)

Limitations of GEDCOM

- ▶ Media storage rather than just linking
- ▶ Current reliance on Notes can bury important information

The beginning of the GEDCOM file

0 HEAD

1 SOUR FTM

2 VERS Family Tree Maker (20.0.0.376)

2 NAME Family Tree Maker for Windows

2 CORP Ancestry.com

3 ADDR 360 W 4800 N

4 CONT Provo, UT 84604

3 PHON (801) 705-7000

1 DEST GED55

1 DATE 07 MAR 2011

1 CHAR ANSI

**1 FILE C:\Users\Owner\Documents\Family Tree
Maker\Tillman_2010_2011-03-07.ged**

1 SUBM @SUBM@

1 GEDC

2 VERS 5.5

2 FORM LINEAGE-LINKED

Continuing the early part of the GEDCOM file

0 @SUBM@ SUBM
1 NAME Hope Nelson Tillman
1 ADDR 6 Saw Mill Brook Way
2 CONT Woburn, MA 01801
2 CONT USA
1 EMAIL hope@hopetillman.com
1 PHON 781-932-8240
0 @I3@ INDI
1 NAME Living /Tillman/
1 SEX F
1 FAMS @F5@
1 FAMS @F3@
1 FAMS @F4@
1 FAMC @F1@

Moving through the GEDCOM file

Individual records (INDI)

- Numbering starts with person whose lineage is being followed
- The individual records (INDI) define
 - My Father (ID 1—@I1@)
 - My Mother (ID 2)
 - My Father's Second Wife (ID 3)
 - ID4 – My Paternal Grandfather
 - ID5 – My Paternal Grandmother
 - ID6-8 – My Paternal Uncles (my Father's siblings)

The family record (FAM)

- links the husband (HUSB), wife (WIFE), & each child (CHIL) by their ID numbers

GEDCOM files for Jonathan Barney my 7th Great Grandfather

Individual Record

0 @I2682@ INDI

1 NAME Jonathan /Barney/
1 SEX M
1 BIRT
2 DATE 29 MAR 1677
2 PLAC Salem, Essex, Massachusetts
1 DEAT
2 DATE NOV 1706
2 PLAC Newport, Newport, Rhode
Island, USA
1 FAMS @F1141@
1 FAMC @F1143@

Family Records

To follow up with his children

0 @F1141@ FAM

1 HUSB @I2682@
1 WIFE @I2683@
1 CHIL @I115@
2 _FREL Natural
2 _MREL Natural
1 MARR
2 DATE 1699
2 PLAC Newport, Newport,
Rhode Island, USA

To get to his parents

0 @F1143@ FAM

1 HUSB @I2687@
1 WIFE @I2688@
1 CHIL @I2682@
2 _FREL Natural
2 _MREL Natural

Importance of GEDCOM

- ▶ **Standard for transmitting data – not storing data**
- ▶ **Enables linking of records according to family lineage and other data relationships**
- ▶ **Links are bidirectional (points to parents, points to offspring) where links in XML are unidirectional**
 - A CHIL tag in the FAM record connects a family to a child
 - A FAMC tag in the INDI record connects a child to a family.
 - HUSB and WIFE tags in the FAM record connect to INDI records, and in the opposite direction, FAMS tags in the INDI record handle both spouses' connection to a FAM record.

XML (Extensible Markup Language)

XSL (Extensible Stylesheet Language)

Definition: XML is a worldwide standard to define data formats for all types of media and for all types of formats (print, web, etc.)

Not yet the standard for genealogy but in discussion. It would allow for improved communication of genealogical data

Purpose:

1. To facilitate the exchange of data in a structured, somewhat meaningful format
2. To allow the processing and presentation of data

What does an XML file look like?

Uses tags with matching end tags, indicates relationships and/or structure by nesting instead of level numbers

```
<indi>  
<id>H0001</id>  
<name>Abraham</name>  
<surname>Howe</surname>  
<sex>M</sex>  
<birt>  
  <date>abt. 1600</date>  
  <town>Hatfield, Broadoak, Essex</town>  
  <country>England</country>  
</birt>  
<family>H1</family>  
</indi>
```

Putting your data on your own website

- ▶ Advantages of using your own website
 - You are free to do it however you want
 - Very accessible and easy to make corrections
- ▶ Disadvantages of using your own website
 - Not likely to live longer than you do
- ▶ Considerations
 - Timetable and plan for adding data after initial gedcom load versus using new gedcom loads
 - How best to display photographs
 - Register with search engines
 - Follow [NGS standards for publishing web pages on the Internet](http://www.ngsgenealogy.org/galleries/Ref_Researching/gswPages.pdf)
(http://www.ngsgenealogy.org/galleries/Ref_Researching/gswPages.pdf)

Putting your data on your own website

Two types of Online Database Managers

- ▶ Gedhtree: <http://www.gedhtree.com>
 - Creates hundreds to thousands of static pages which will then be uploaded onto a website for viewing.
 - Harder to maintain
 - Easily indexed by Google and other search engines
- ▶ The Next Generation of Genealogy Sitebuilding: <http://lythgoes.net/genealogy/software.php>
 - Creates web pages on the fly from a web-enabled database
 - Easier to maintain
 - Easily indexed by Google and other search engines
- ▶ For more information see Cyndi's list: <http://www.cyndislist.com/software.htm#HTML>

Repositories: Ancestry

- ▶ World's largest for-profit company in this marketplace
- ▶ Has purchased much of its competition
- ▶ Links between its many products and websites: Family Tree Maker, Ancestry.com, Genealogy.com, MyFamily.com, Rootsweb, etc.

Repositories: FamilySearch.org

- ▶ Go to this afternoon's talk
- ▶ FamilySearch.org is a free website provided by the Church of Jesus Christ of Latter-day Saints.
- ▶ LDS has a long history with providing genealogical resources
- ▶ First came online in 1999.


Repositories: Social Publishing/Social Networking

- ▶ Rootsweb
- ▶ Familylink, Geni (Facebook),
- ▶ Scribd.com \$, Lulu.com \$
- ▶ Vanity publishing – no external validation
- ▶ Crowdsourcing, self-policing as feedback mechanism (may or may not be expert validation)
- ▶ Beware as there will be a market shakeout of these in the next few years and your choice may or may not survive

How Your Choices of Where You Put Your Data Meet Your Need

	Accessibility	Permanence	Correction Handling
Your own website	High	Questionable beyond your lifetime	Good. Ease depends on software choice
Ancestry	\$	Presumably high	Periodic updating possible
FamilySearch	High	Presumably high	Corrections can be added. Old will remain
Social publishing	Questionable	Questionable	Questionable

Evolution of information across media


- ▶ **Paper is the most long-lasting/not the most accessible.**
 - ▶ **Do not preserve online solely to get rid of paper; all media complement one another. Remember: LOCKSS (lots of copies keep stuff safe)**
 - ▶ **Micro formats were an early part of format evolution, universally hated BUT long lasting as compared to digital formats.**
- 

Evolution of information across media

- ▶ **All media formats need to be refreshed periodically or moved to new format when equipment no longer available.**
 - **Start by scan of paper/micro format to put into electronic form**
- ▶ **Need to rescan as media evolves**
 - Example: rescanned original census records for ease of use**
- ▶ **Overarching concerns for all media:**
 - **Standardization and Sourcing**

Digital formats

- ▶ Photo: .jpg, .gif, .png, .tif, etc.
 - ▶ Audio: .wav, .au, .mp3, etc.
 - ▶ Video: .mov, .mpeg, etc.

 - ▶ Quality versus compression
- 

Evolution of Information Across Media

Photographs

- ▶ **Resolution** – print resolution took too long to load on early web browsers (Vatican Scrolls)
 - Digital cameras – The low resolution output of early digital cameras not seen as acceptable now.
 - Early scanning, photocopying techniques
- ▶ Black and white, Sepia, Color, Hand-colored
- ▶ How do you create online wall of family photographs?
 - Scan to improve quality of existing photographs
 - Scan to share with others
 - Scan to preserve

Photographs

- ▶ Reason to keep at least two versions of each image
 - Original high resolution
 - Edited versions
 - Restoration of faded image to bright new looking
 - Small version for fast loading on web
 - Cropped version – thumbnails for quick look backed up by higher resolution

- ▶ What do you need to know to manage your images
 - Lossy versus Not lossy
 - Software tricks

Evolution of Information Across Media

Audio/Video

- ▶ See [Wikipedia table of audio formats from 1877 to date](#)

- ▶ Compression

- ▶ Abundance

Now so much easier to produce/keep – it needs to be married to rest of records by metadata.


- ▶ Preservation of audio/video: online, CDs, DVDs – quickly can become a space and refreshing issue.

Lessons learned: history of storage digital formats

- ▶ Formats keep changing and no stability is expected in the future
 - Punched cards, Magnetic tape
- ▶ Capacity grows and size shrinks
 - Hard disk drives
 - Floppy disk shrank to 5 shrank to 3 went away
 - Memory devices (zip drives, flash drives)
- ▶ Cloud as a more recent development --an outgrowth of the changes/developments in networking. What's next?
- ▶ Importance in refreshing media regularly to keep up with changes and because the media itself may become unusable
- ▶ You need to plan ahead: **What is your own personal preservation plan?**

**Those who don't know history are destined to repeat it
(Edmund Burke (1729-1797))**


Personal Preservation Plan

1. Make a plan.
 2. Associate metadata with your data.
 3. Identify the data you need to backup
 4. Use multiple approaches and make multiple copies
 5. Plan for addition of data elements to save
 6. Plan for the transition of digital data to new storage media.
 7. Plan for transitions in data stewardship: pick trustworthy repositories.
 8. Take into consideration privacy and security
 9. Review and update plan regularly
- 

Don't forget about copyright

- ▶ Just because you “can” copy something does not make it legal to do so.
- ▶ Importance of keeping permission and source information/
- ▶ Data is not copyrightable, but your words are.
- ▶ Copyleft: what rights do you want to allow?
 - How much protection do you want to retain if your goal is to make the info found by future generations?

Stay tuned

- ▶ Answers keep changing
 - ▶ Standards evolve
 - How will XML play with GEDCOM?
 - Will something else develop?
 - ▶ Media
 - Changes in media will need to be reflected
 - ▶ Players
 - How will social media impact marketplace?
 - Will Ancestry and FamilySearch continue as the big players?
- 

Stay tuned

- ▶ Other possible considerations
 - DNA integration - sampling today too small but growing by leaps and bounds
 - How will you point future generations to the family videos and youtubes of today?

Q&A

Feel free to contact us.

We belong to both MSOG and ESOG.

Hope N. Tillman

<http://www.hopetillman.com/genealogy>

hope@hopetillman.com

Walt Howe

<http://www.howegenealogy.org>

walt@walthowe.com