Preserving Genealogy Data Online

Hope N. Tillman & Walt Howe
Daughters of the American Revolution
October 15, 2011

- We want the family history 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

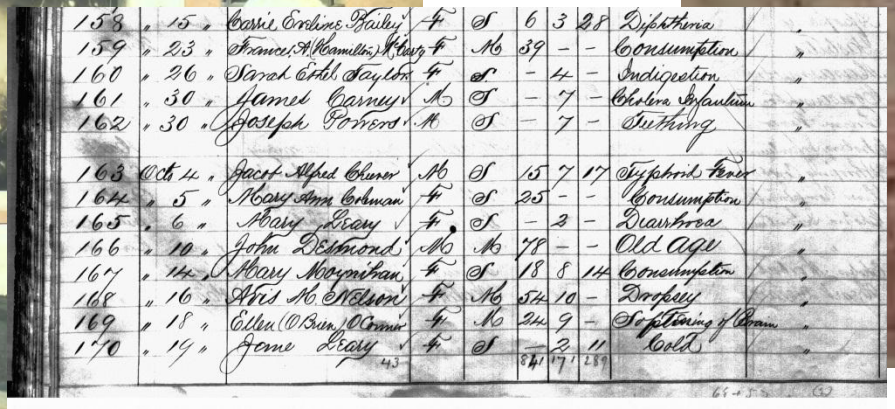
What we will be covering

- Personal preservation plan
- Your choice of genealogy software
- GEDCOM
- XML
- Putting your data on your own website
- Repositories
- Evolution of information across media

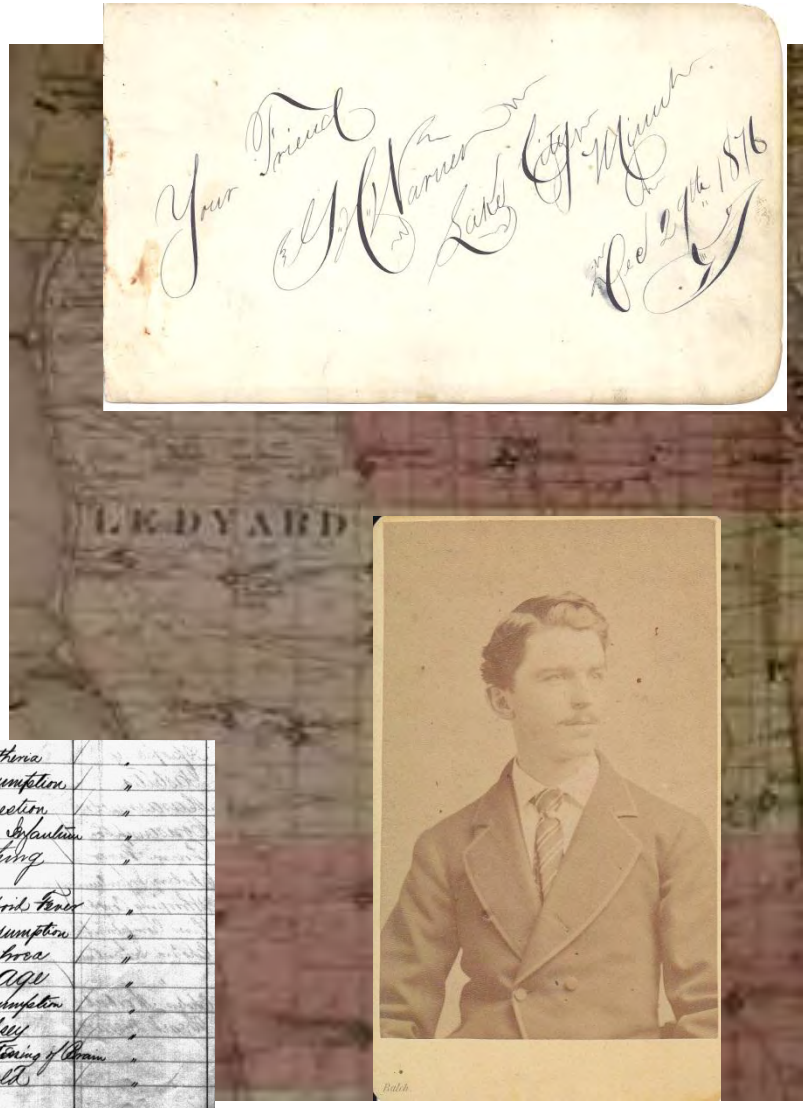
You need a Personal Preservation Plan!

1. Make a plan

   Ask yourself what you want to save and why?



158	"	15	"	Carrie Evelina Bailey	F	S	6	3	28	Diphtheria
159	"	23	"	Frances A (Hamilton) Kitz	F	M	39	–	–	Consumption
160	"	26	"	Sarah Estb. Taylor	F	S	–	4	–	Indigestion
161	"	30	"	James Carney	M	S	–	7	–	Cholera Infantum
162	"	30	"	Joseph Powers	M	S	–	7	–	Teething
163	Oct	4	"	Jacob Alfred Ebner	M	S	15	7	17	Typhoid Fever
164	"	5	"	Mary Ann Coleman	F	S	25	–	–	Consumption
165	"	6	"	Mary Leary	F	S	–	3	–	Diarrhoea
166	"	10	"	John Desmond	M	M	78	–	–	Old Age
167	"	14	"	Mary Moynihan	F	S	18	8	14	Consumption
168	"	16	"	Alex M. Nelson	F	M	54	10	–	Dropsy
169	"	18	"	Ellen O'Brien O'Connor	F	M	24	9	–	Softening of Brain
170	"	19	"	Jane Leary	F	S	–	2	11	Cold



Personal Preservation Plan

1. Make a plan
2. Associate metadata with your data

   Metadata is information describing data

   An image may include metadata that describes who or what it is, 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 original source document is located, and a short summary of the document.

   Tag documents, images, etc. appropriately and consistently with metadata chosen: surname, date, location, relationship, or 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 – where, what, when, why
    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 –
   - **For now online:** 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 online:** Ancestry, FamilySearch
   - **Long term offline**: Historical Societies, Libraries, the D.A.R. Library, Family Members

LOCKSS - Lots of Copies Keep Stuff Safe

Follow the LOCKSS model (lockss.stanford.edu)

- Keep originals AND **REGULARLY**
- Backup to flash drives
- Backup to external HDD
- Back up to cloud service

Cloud Backup

Features to consider: platform supported, encryption, type of plan, maximum file size, syncing, auto-detecting changes.

Current major players:

- Amazon Cloud Services: www.amazon.com
- Apple iCloud: www.apple.com/icloud/
- Backupify: www.backupify.com
- Box: www.box.net
- Carbonite: www.carbonite.com
- Dropbox: www.dropbox.com
- Google Drive (coming soon)
- Idrive: www.idrive.com
- Mozy: www.mozy.com
- Sugarsync: www.sugarsync.com
- Syncplicity: www.syncplicity.com/
- Windows Live Skydrive: skydrive.live.com

Market shakeout continues

Linking to External Documents versus Creating Copies of External Documents

- **Consider whether you want to**
  - To link to external documents or
  - To create copies of external documents or
  - Whether you should do both:
- **Questions to consider**
  - 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. Save regularly. Don't wait until you are done because that will never happen.

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. Save regularly. Don't wait until you are done because that will never happen.
6. 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. Save regularly. Don't wait until you are done because that will never happen.
6. Plan for addition of data elements to save
7. 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. Save regularly. Don't wait until you are done because that will never happen.
6. Plan for addition of data elements to save
7. Plan for the transition of digital data to new storage media.
8. 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. Save regularly. Don't wait until you are done because that will never happen.
6. Plan for addition of data elements to save
7. Plan for the transition of digital data to new storage media.
8. Plan for transitions in data stewardship: pick trustworthy repositories.
9. 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. Save regularly. Don't wait until you are done because that will never happen.
6. Plan for addition of data elements to save
7. Plan for the transition of digital data to new storage media.
8. Plan for transitions in data stewardship: pick trustworthy repositories.
9. Take into consideration privacy and security
10. 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

Genealogy Software Choices

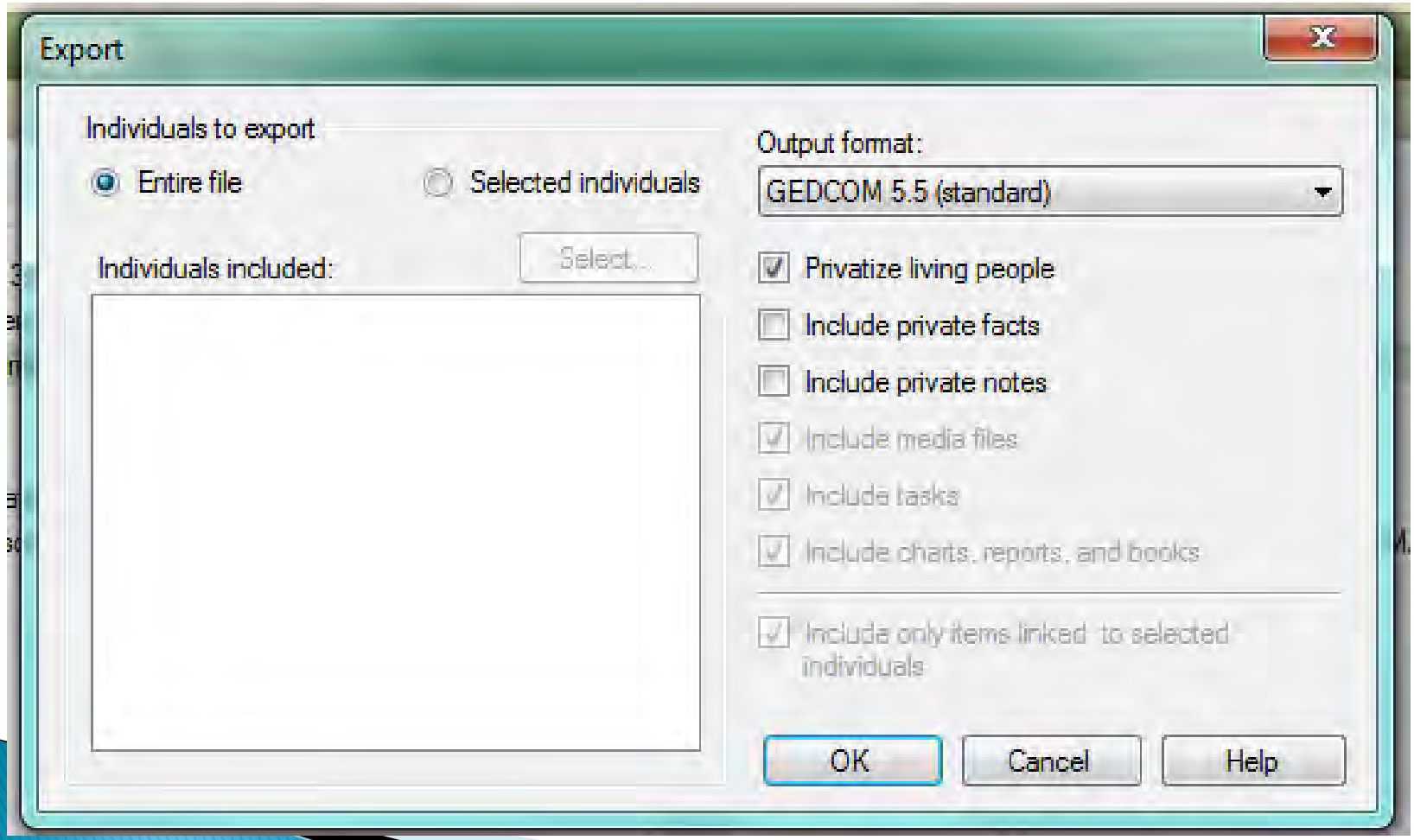
- Choose a package with a comfortable interface for you
- Decide what kind of sharing is right for you?
- Update your software regularly
- Be prepared to change your software, as companies are acquired or leave the marketplace. Market shakeout continues.

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. 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. 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

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 – But it can!
- 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 (www.ngsgenealogy.org/galleries/Ref_Researching/gswebPages.pdf)

Putting your data on your own website

Two types of Online Database Managers (start w/GEDCOM)

- Gedhtree: 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: 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
  - Let her know about any dead links (hard to keep up with)

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, Footnote.com, Genealogy.com, MyFamily.com, Rootsweb, etc.

Repositories: FamilySearch.org

- 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: DAR Library

- Premier collection with excellent website
  - Located in Washington, D.C.
  - Unique collection of material available nowhere else
  - Plus over 185,000 monographic volumes, 300,000 digitized files, 65,000 microforms
  - Current project of digitizing member application papers
- Finding tool: *American Genealogical Research at the D.A.R.*
- This group is much more knowledgeable about these resources than we are.

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
- These sites may put you in touch with family members you may never have reached otherwise!

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	Updating capabilities improving
FamilySearch	High	Presumably high	Corrections can be added. Old will remain
Social publishing	Questionable	Questionable	Questionable Variable

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 (i.e. microfilm and microfiche) 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 greater readability and ease of use
- Overarching concerns for all media:
  - Standardization and Sourcing are very important.

Evolution of Information Across Media

Text

- OCR – optical character recognition
  - Becomes indexable and searchable
- Character sets –
  - **ASCII** - American Standard Code for Information Interchange (original 26 letters, 10 numbers and printer controls – keyboard characters)
  - **ANSEL** – American National Standard for Extended Latin Character Coded Character Set
  - **Unicode** - is a worldwide standard for the consistent representation and handling of text expressed in most of the world's writing systems.

Digital formats

- Photo: .jpg, .gif, .png, .tif, etc.
- Audio: .wav, .au, .mp3, etc.
- Video: .wmv, .mov, .mpeg, .mts, .asf, .flw, etc.

- Many formats
- You may need to know the best one for your purpose
- 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

- See list of photo editing software products on your handout
- 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 (low resolution) for fast loading on web
    - Thumbnail version for quick look backed up by higher resolution
- What do you need to know to manage your images?
  - Lossy (does file lose detail when saved?)
  - Software tricks

Evolution of Information Across Media
Audio/Video

- See Wikipedia table of audio formats from 1877 to date
- Compression versus quality
- Abundance
  - Explosion of audio video today.
  - Now so much easier to produce/keep – it needs to be married to rest of genealogical 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. Save regularly. Don't wait until you are done because that will never happen.
6. Plan for addition of data elements to save
7. Plan for the transition of digital data to new storage media.
8. Plan for transitions in data stewardship: pick trustworthy repositories.
9. Take into consideration privacy and security
10. 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 getting and keeping permission and source information
- Data is not copyrightable, but your words are.
- 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?
  - You can embed what you want to allow using Creative Commons copyright licenses.

Stay tuned

- Answers keep changing
- Standards evolve
  - How will XML play with GEDCOM?
  - Will something else develop?
- Changes in media will continue to need to be considered.
- Players
  - Who will be the big players going forward?
  - How will social media impact the marketplace?

Stay tuned

- Other possible considerations
  - Rebranding: We still refer to New England History Genealogical Society as NEHGS but they are rebranding themselves as American Ancestors with their website as AmericanAncestors.org
  - Popularization of family history by "Faces of America" and "Who Do You Think You Are?"
  - How will you point future generations to the family videos and YouTubes of today?
  - DNA integration - sampling today too small but growing by leaps and bounds

Q&A

Thank you for attending

Hope N. Tillman
http://www.tillmangenealogy.org
hope@hopetillman.com

Walt Howe
http://www.howegenealogy.org
walt@walthowe.com