

GEDCOM Files



Table of Contents

	0
Part I Introduction	4
1 Overview of Terms Used.....	4
Part II How GEDCOM Works	7
1 Source Tags.....	7
2 Compatibility Issues.....	10
Part III Computer Encoding	12
1 Issues with Character Sets.....	12
2 Date Format.....	13
Part IV Testing Import and Export	15
1 Tags Tested.....	15
2 Programs Tested.....	15
Test Results	17
Part V Summary	20
1 About the Document.....	21
2 About the Author.....	21

Part



1 Introduction

When you store your family history information with a software program you may realize overtime the features your current program has is not what you are looking for. So you decide to try a new program that you think will meet your needs. It is like buying a new cell phone. You look at all the features each phone has then upgrade your phone and cell plan as easy as 1, 2, 3. You don't have to worry about losing your data (unless the phone breaks and it is not synced to a computer) or worry that you can't make a phone call to someone who uses a different carrier. Today, if you decide to change to a different carrier, you don't have to worry about losing your phone number.

This should apply to genealogy programs as well, if you want to try a different program, you should not have to worry about losing your data, and you shouldn't have to add all your documents again. But this is not the case. So what causes this to happen? The main problems are the way programs are written to store, read, and write your genealogy information. To understand the issues in detail, we will explore the GEDCOM standard with examples, and do a comparison of several programs so you can have a better understanding.

What is GEDCOM

GEDCOM is an acronym for Genealogical Data COMMunication. GEDCOM was developed by the Church of Jesus Christ of Latter-day Saints. The current release is 5.5. The GEDCOM format was created to help standardize the ability to export and import files from one genealogy application into another. To understand how this is accomplished, some terms need to be explained.

1.1 Overview of Terms Used

GEDCOM: Provides a standard for programmers to create genealogy programs that read and write family history information. This is accomplished through its ability to store data in "fields" that are stored with "tags". This is similar to spreadsheet that has rows and columns. The header is the tag, and the information stored in the cells is the field.

Tags: are generally four (4) characters in length and are mostly predefined. Custom tags can also be created. Tags allow a field to occur any number of times within a record. Tags are set in a hierarchy that is used to capture and store the data associated for people's names, events, sources, notes, media, etc. The concept is similar to storing information in a spreadsheet.

	A	B
1		Books Sold
2	January	200
3	February	110
4	March	175

Hierarchy: The beginning of a new logical record is designated by a line whose level number is 0 (zero).

Each new level number must be no higher than the previous line plus 1.

Fields: Stores the data for each tag. The programmer defines how many characters can be stored in a field.

Records: The field and tags that belong together are records.

Computer Encoding: Is how a program reads and writes the characters (A-Z, numbers, etc.) stored on a computer.

Person Information A Program Screen Used to Input Data:

Tags:

GIVN Given Name * Henry

SURN Sur/Maiden Name: Reece

NICK Nick Name: Buddy

Gender * Male

Individual Events

Event Name	Date	Place	Det
Birth	25 Dec 1810	Dalton, Whitfield, Georgia, USA	
Death	06 Nov 1885	Big Flat, Baxter, Arkansas, USA	
Burial		Big Flat, Baxter, Arkansas, USA	

Fields: used to input /stores information

Hierarchy:

0 0000 INDI

1 0001 NAME Henry /Reece/

2 0002 SURN Reece

2 0003 GIVN Henry

2 0004 NICK Buddy

GEDCOM File:

Fields:

1 BIRT

2 DATE 25 Dec 1810

2 PLAC Dalton, Whitfield, Georgia, USA

1 DEAT

2 DATE 06 Nov 1885

2 PLAC Big Flat, Baxter, Arkansas, USA

1 BURI

2 PLAC Big Flat, Baxter, Arkansas, USA

Tags:

Screen Overview:

Although you see "Given Name" as part of the screen where you input your data, the tag that actually stores the information is GIVN.

Tags are created in a hierarchy so information stays together. SURN, GIVN, and NICK are subsets of NAME. DATE and PLAC are subsets of birth, and death.

Part



2 How GEDCOM Works

In order for information to transfer from one program to another, the tags need to be readable and writeable. GEDCOM has many tags that can be found in the GEDCOM 5.5 Standards Document that are used to store the various pieces of a person's family history.

Compatibility Issues Due to Tags

So the concept of following a standard seems simple enough and the data from one program should transfer seamlessly to another, right? Well, that is not necessary what happens. Programs do not all work the same and information may not be transferred from one program into another.

There are many GEDCOM tags that are used to store genealogy information. A company may decide that some tags are not important to them, so they exclude them.

They can also manipulate tags in three (3) ways. When importing a GEDCOM file from one program to another they can write executable changes to tags where they combine two tags into one upon importing the information into their program. They can also change a tag name, or decide which sublevel belongs to which tag.

They can also create their own custom tags that are not readable in other programs. This is mainly seen with source citation templates, and the extra information that they store. Individual and marriage events are allowed to be custom as per the GEDCOM standard, and most programs do read and write them. Below are the standard individual events:

Adoption	Adult Christening	Bar Mitzvah	Bat Mitzvah	Birth
Blessing	Burial	Caste	Cause of Death	Census
Christening	Confirmation	Cremation	Death	Education
Emigration	First Communion	Graduation	Immigration	Naturalization
Occupation	Ordinance	Ordination	Phone Number	Physical Description
Probate	Property	Religion	Residence	Retirement
Social Security Number	Title	Will		

They can also decide which tags their programs will read and write. A tag can be read by a program, but may not write the information, so when you import your information, it will not be transferred.

Upon exporting a GEDCOM file, a program can decide not to write all the tags, even if they are common tags, and therefore your information will be lost.

See Programs Tested to get results on how various companies have omitted and manipulated tags.

2.1 Source Tags

Below is an example of tags with sources which are a bit more complicated than GIVN, NICK, BIRT, DATE, PLAC, which are common tags for all people in your file.

It becomes a bit more complicated with sources are used to back up the information in a family tree by adding the evidence that supports events, dates, places, and relationships.

When it comes to programming and saving information for sources, there are four (4) elements: the Repository Record, Source Repository Citation, Source Record, and Source Citation. Some of the tags share a common sub level tag, such as NOTE and TEXT.

REPOSITORY_RECORD

Is used to capture the information of where the data was found. Uses tags such as: NAME, ADDR, NOTE.

SOURCE_REPOSITORY_CITATION

Is used for information such as call number, media attached, or notes. Uses tags such as: CALN, MEDI, NOTE.

SOURCE_RECORD

The SOURCE_RECORD structure was simplified into five basic sections: author, title, publication facts, and repository. Uses tags such as: EVEN, PLAC, AUTH, TITL, ABBR.

SOURCE_CITATION

The source citation contains only information specific to the fact being cited and the quality of the source. Uses tags such as: EVEN, TEXT, PAGE


```
1 SOUR @S4@
2 PAGE 1860 Federal Census, Bigflat, Searcy, Arkansas; Taken 16 Jun 1860;
3 CONT Page 22; Dwelling # 142
2 DATA
3 TEXT Dwelling # 140 James Sutterfield Dwelling # 141 A J Sisk Dwelling
4 CONT #142 Henry Reece
2 NOTE Henry is 50 yrs old
```

1

Family Tree Maker

```
2 SOUR @S00004@
3 PAGE 1860 Federal Census, Bigflat, Searcy, Arkansas; Taken 16 Jun 1860;
4 CONC Page 22; Dwelling # 142
3 DATA
4 TEXT Dwelling # 140 James Sutterfield Dwelling # 141 A J Sisk Dwelling #142
5 CONC Henry Reece
```

2

Family Tree Maker

```
0 @N00002@ NOTE
1 CONC Henry is 50 yrs old
```

3

```
1 SOUR @S4@
2 PAGE 1860 Federal Census, Bigflat, Searcy, Arkansas; Taken 16 Jun 1860;Page
3 CONT 22; Dwelling # 142
2 DATA
3 TEXT Dwelling # 140 James Sutterfield Dwelling # 141 A J Sisk Dwelling
4 CONT #142Henry Reece
1 SOUR @S7@
```

4

Screen Overview:

As stated before, tags are created in a hierarchy so information stays together. When the hierarchy is not followed, the information can be lost.

Each level plays a role in supporting the main tag. Looking at "NOTE" in the first image above, this is a sub tag under SOUR. Then looking at the next two (2) images, Family Tree Maker took the tag and changed it into a NOTE for the person. This is apparent because it is no longer associated with the SOUR (as seen in image two (2)). In image three (3), it is assigned level 1 and turned into a NOTE for the person. So when the information was imported back into the original program, the source note was lost, as seen in image four (4).

2.2 Compatibility Issues

Each tag has an area to add information “in a field”. Each field has a defined number of characters that are allowed in the field. As an example: The field PUBLication is used to store information about a book, microfilm, etc. If the length of the field in the program you are using is longer than the length of the field you transfer your information to, the information will be shortened if your data contains more characters than allowed by the new program.

How to Fix Tag & Field Compatibility Issues

Information would transfer properly from one program to another without data loss if all the defined tags were considered mandatory, followed the proper hierarchy, didn't change tags during an import process, and every field had a defined length that was adhered to. However, it is highly unlikely that all programs will ever agree to have all tags considered mandatory, so maybe a subset of tags should be considered mandatory, or a list of the tags used by a program should be outlined so a potential customer would know if some of their information has the potential of being lost before trying a new program.

Part



3 Computer Encoding

Computer encoding is used for reading and writing everything stored on a computer. The ANSEL, ASCII, is a plain text character set, where as Unicode and UTF-8 support international character sets.

Computer Encoding Incompatible Example

```
1 PUBL State of Arkansas, Department of Health, Division of vital Records,
2 CONT "Certificate of Death"

1 PUBL st̄ate ōf Arkansas, Department of Health, Division of vital
2 CONC Records, aa? certificate of Death aa??
```

Screen Overview:

A file created using UTF-8 with quotes, was imported into a program that saved the information using ANSEL, and turned the quotes into gibberish.

3.1 Issues with Character Sets

A programmer can decide what characters are allowed in each and every field so errors can occur at the overall programming level and at the field level. If your program allows for a slash or quotes in a field that the program you are transferring your information to (even if both are using UTF-8), your data can be lost or messed up, or not import at all.

This is an issue especially with the NAME tag, because GEDCOM stores a slash to separate the GIVN and SURN. If an extra slash or quotes are allowed in NAME, the import can fail as the program may not know what to do with the extra slash.

```
0 @I1@ INDI
1 NAME Henry /Reece/
2 GIVN Henry
2 SURN Reece 1

0 @I00001@ INDI
1 NAME Henry /Reece/Gray/ 2

0 @I00001@ INDI
1 NAME Henry "Buddy" /Reece/ 3
```

Screen Overview:

1. The first program follows the correct hierarchy for name which is important especially when quotes or slashes are used.
2. The second example uses a slash to distinguish a person with 2 last names due to adoption. A program that does not use the 2nd level (GIVN and SURN), may not know how

to split the name.

- The third example shows a program that does not allow for nick names, so data was added in quotes. If the user imports this file into another program, the information may not import properly if quotes are not allowed. If the new program has a field for nick name, you will have to decide to continue to add nick name to the name field.

3.2 Date Format

Another issue that can arise is how a programmer programs the ability to import and export dates. There are a couple of ways a date can be written. Month, day year; day, month year; with various ways of adding by using slashes, dashes, or just spaces. Then some programs allow abt, bef, aft, and bet. With the combination of ways to enter in dates, it is possible you can lose date information if the program you change to does not support a date format used in another application. It is also possible the data will not import at all because it doesn't know what to do with your data because it can't be read by the application.

The image shows a dialog box titled "Set your date format". It contains two main sections, each with a radio button and three sub-options. The first section is "Day Month Year" (selected) with options: "dd/MM/yyyy Ex: 11/12/2008", "dd MMM yyyy Ex: 11 Dec 2008" (selected), and "dd-MM-yyyy Ex: 11-12-2008". The second section is "Month Day Year" with options: "MM/dd/yyyy Ex: 12/11/2008", "MMM dd yyyy Ex: Dec 11 2008", and "MM-dd-yyyy Ex: 12-11-2008". An "Update" button is located at the bottom right.

Part



4 Testing Import and Export

A sample file with one (1) person was used to see how the information imported, was read, and exported.

Tags for nick name, pre and suffix, alias, givn, and surn were not tested because there is too much of a disparity between the programs of which tags are used.

Notes for the person were tested. For the most part notes were added correctly. Some minor issues with CONT and CONC and ability to read UTF-8 characters were encountered.

Nine (9) Events were tested. Six (6) were standard, and three (3) were custom created.

Three (3) Source Citation tags and six (6) Repository tags were tested.

Templates of how sources are entered and stored were not part of the testing because there is no template standard that has been adopted across all genealogy programs.

Rating tags were not tested as rating is not universal across all programs and is not part of the GEDCOM standard.

4.1 Tags Tested

NAME – Name

GIVN – The majority of the program did not use this tag and was excluded from the results.

SURN – The majority of the program did not use this tag and was excluded from the results.

EVEN – Standard Events: BIRT, BURI, RESI, CENS, DEAT, OCCU,
Custom Events start with EVEN, followed by level 2 TYPE.

SOUR – Source tag has the following sub tags: PAGE, TEXT, NOTE

REPO – Repository has the following sub tags: TITL, AUTH, PUBL, NAME, NOTE, TEXT

NOTE – Level 1 tag for notes about a person

4.2 Programs Tested

A group of programs were tested and the outcome is not to prove one application is better than another. It is merely to show there are differences with programs and help genealogists understand that they may lose data going from one application to another.

Some programs that were considered, but omitted from the testing because they do not support import, or export of GEDCOM files or do not support sources are Geni, We Relate, Familylink, and FamilyBuilder. Considered but omitted because they required a subscription fee: OneGreatFamily, and FindmyPast.

There are numerous genealogy programs to test, and a list can be found at: <http://www.gensoftreviews.com/>

Internet Based Programs

	Ancestry.com	MyHeritage	ourFamilyology	Tribal Pages
Events	ok	ok	ok	see notes
CITATION Details SOUR Tags:				
PAGE	ok	see notes	ok	n/a
TEXT	ok	see notes	ok	n/a
NOTE	ok	n/a	ok	n/a
Source Information REPO Tags:				
NAME	shortened	n/a	ok	n/a
TITL	ok	see notes	ok	ok
AUTH	ok	n/a	ok	see notes
PUBL	ok	n/a	ok	see notes
NOTE	see notes	n/a	ok	n/a
TEXT	see notes	n/a	ok	see notes

PC Based Programs

	AgeLong Tree	Ancestral Quest	Family Historian	Family Tree Legends	FamilyTree Maker
Events	see notes	ok	ok	ok	ok
CITATION Details SOUR Tags:					
PAGE	n/a	ok	ok	ok	ok
TEXT	n/a	ok	ok	ok	ok
NOTE	n/a	ok	ok	ok	ok
Source Information REPO Tags:					
NAME	n/a	shortened	n/a	shortened	ok
TITL	see notes	ok	n/a	ok	ok
AUTH	n/a	ok	ok	ok	ok
PUBL	n/a	ok	ok	ok	ok
NOTE	n/a	ok	ok	ok	ok
TEXT	shortened	ok	ok	ok	n/a
ADDED:		@R ABBR (copied TITL info)			@R ADDR (blank) @R Email @R Phon

	GenBox	Legacy 7	The Master Genealogist	Roots Magic 4	Reunion
Event	ok	see notes	see notes	ok	ok
CITATION Details SOUR Tags:					
PAGE	ok	ok	ok	n/a	see notes
TEXT	ok	ok	n/a	see notes	see notes
NOTE	ok	ok	n/a	n/a	see notes
Source Information REPO Tags:					
NAME	n/a	shortened	see note	shortened	ok
TITL	ok	ok	ok	see notes	ok
AUTH	ok	ok	ok	n/a	ok
PUBL	ok	ok	ok	n/a	ok
NOTE	ok	ok	see note	n/a	see notes
TEXT	ok	ok	see note	n/a	see notes
ADDED:	@S _TYPE	@R ADDR (copied NAME info) ABBR	ABBR (added part of TITL)	@SOUR_TMPTL FIELD NAME (Page)	

Screen Overview

Ok indicated the information transferred from the import and export successfully.

n/a indicates the information did not transfer properly.

Shortened indicates the field does not have enough space to accommodate the original text.

Added indicates tags were added during the export process from the imported program vs. the original data.

See note indicates further information about the tag was captured. Read the test results in the notes section.

4.2.1 Test Results

Ancestry.com – Sources not attached to events do not show up under the person's timeline.

REPO: AUTH, PUBL, NOTE, and TEXT do not show up anywhere in the program, but the information did import as it showed up in the export with the caveat that the NOTE and TEXT was combined into 1 tag.

AgeLong Tree – Converted @R to @S. Custom events did not import or export.

Family Historian – Converted @R to @S

FamilyTreeMaker – Sources that are not attached to events for a person are not shown in the family view. They are shown in the Source Group Tab.

Legacy – Due to its Sourcing incompatibility, the file did not import correctly. The person's name and information did not appear in the program. But the file did read information and exported. The program added sentence structure to Events adding unnecessary data.

MyHeritage – Only has TITL, PAGE, and TEXT tags for sources. Was not able to test the export, requires a backup subscription fee.

Reunion – Combined SOUR tags: PAGE, TEXT, and NOTE to PAGE. Combined REPO tags: NOTE and TEXT to TEXT.

RootsMagic – Combined parts of the original NAME, TITL tags. It also combined SOUR Page and Text to 1 field. The program added standard events to each person which added an extra 22k per person to the exported GEDCOM file. Sources not added to events are not shown for the person.

The Master Genealogists – Combined NOTE and TEXT into 1 field - NOTE. Custom events were converted into notes and misc events.

Tribal Pages – Sources are not attached to people or events. Import Titles only. Program tags include: TITL, ABBR, AUTH, PUBL, TEXT, URL, TYPE, but the only tag that exported was TITL. You can add custom events, but only birth, marriage, and death events exported. The note added several lines of gibberish (ANSEL coding is used).

How to Create your own Test

1. Export a GEDCOM file from your current program.

2. Import that file into another program.
3. Save the file and export the file as a GEDCOM.
4. Open both files using notepad and compare the results.

Part



5 Summary

There are some shortcomings of the GEDCOM standard that need updating and changes to create better family trees. Some things that need to be considered, but not limited to:

A portion of tags needs to be made mandatory so they will transfer information from one program to another. Tags being used in each program should be listed so customers can decide which program is best for them.

The character set that a program uses needs to be known, so customers know if they will lose data due to unsupported characters that do not transfer.

Fields need to be consistent in length so information is not lost.

Images of all kinds need to be supported in all genealogy programs and transfer to each program.

Evidence needs to be incorporated into family trees better than just the QUAY tag, so the reliability of the information is more evident.

A new standard that addresses some of the newer genealogy research and scientific efforts needs to be created and adopted by all software vendors. fhiso is the organization that is taking on this task. <http://fhiso.org/>

Read more about the GEDCOM standard at:

<http://homepages.rootsweb.ancestry.com/~pmcbride/gedcom/55gcch1.htm#S1>

http://www.ngsgenealogy.org/cs/GenTech_Projects

<http://www.genealogymedia.com/2011/02/14/rootstech-2011-data-model/>

<http://members.cox.net/leebr/genealogy/>

<http://wiki-en.genealogy.net/GDM>

http://en.wikipedia.org/wiki/GEDCOM#GEDCOM_model

Read about some better GEDCOM initiatives that are being discussed and seeking help:

<https://bettergedcom.wikispaces.com/>

5.1 About the Document

This document was created by ourFamilyology, Inc., Franklin Lakes, NJ 07417

©Copyright 2009 All Rights reserved.

Information in this document is subject to change without notice. *Last updated August 16, 2012*

No part of this document may be reproduced in any form without the express written permission of ourFamilyology, Inc.

ourFamilyology® is a registered trademark of ourFamilyology, Inc.

5.2 About the Author

Brandy Sacco is the owner and founder of ourFamily•ology, Inc. as well as the ourFamily•ology online genealogy program. As a genealogy hobbyist with more than 18 years of experience, she recognized the epidemic of people copying family histories that are unverified and not supported up with sources. She felt compelled to remedy this by creating a program that actually helps people create source verified, responsible family histories.

Prior to starting ourFamily•ology, she was an executive at a Fortune 500 company. Her abilities as a team leader were quickly recognized as she spearheaded the development and deployment of a multitude of world-wide projects. She is known for her troubleshooting abilities and has the know-how and can-do attitude to get the job done in difficult to near impossible situations.

Brandy has spoken at RootsTech, is a member of National Genealogical Society, The Bergen County Historical Society, and is involved with several groups that are trying to increase the awareness and advocate standards that will help shape the future of genealogy.

For more information

Please visit <http://www.Family-Genealogy.com>