

The Wayback Machine - <http://web.archive.org/web/20000409193831/http://www...>

GEDCOM 101

by [Jan McClintock](#)

As genealogists use it, the GEDCOM format is famous as a great way to share data, but is infamous for being finicky. In this article I'll attempt to explain the basic principles and formats of a GEDCOM file. For more details (*many* more), you can read the actual GEDCOM v5.5 documentation [here](#). You will find the many intricacies and intriguing potential of this format.

What is GEDCOM?

GEDCOM is an acronym for **G**enealogical **D**ata **C**ommunication; it's a method of formatting the text of your family data so that different software programs and operating systems can read and understand it.

The GEDCOM "standard" was originally developed by the Church of Jesus Christ of Latter-day Saints' Family History Department, which holds the copyright. Although it began as a "universal" format, many genealogy programs implement their own, slightly-modified version of GEDCOM. To you, the user, this means that it isn't always a sure thing that all of your data makes it from one program to another (more about this later).

What is a GEDCOM File?

A GEDCOM file is a series of text lines (usually using the ASCII character set) each holding a specific piece of data relevant to your family file. The lines are numbered to show the hierarchy among the data, and are tagged to show the type of data.

It's quite possible to construct a GEDCOM file manually (using a word processor), but the process could be painstaking. For instance, the GEDCOM of one of my family files in Reunion including only 405 individuals (288 families) ended up at 5306 lines! Reunion will create a GEDCOM file for you, properly formatted and ready to be shared and used by others. Details and instructions are in your electronic manual [[Help -> Search for Help "gedcom"](#)]

Parts of a GEDCOM File

If you created a GEDCOM file and opened it using your word processor, you would see line after line of numbers, abbreviations, and bits and pieces of data. There are no blank lines and no indentations in a GEDCOM file. Although it might look confusing, there is method to the madness.

Groups of lines that hold information about one individual (INDI), one source (SOUR) or one family (FAM) are considered **records**, and each line in a record has a **level number**. The first line of every record is numbered "0" (zero), to show that it is the beginning of a new record.

Within that record, different level numbers pertain to the next nearest level above it; the lines following the 0 level all pertain to that record, until the next 0 level is reached.

Think of it like an outline format, with level numbers instead of Roman numerals, etc. For example, here is a simplified record of an individual with some explanations in parenthesis:

```
0 NAME Joseph /PRYOR/
1 SEX M (data about the record in level 0 above, which is Joseph)
1 BIRT (more data about Joseph, this time that he had a birth...)
2 DATE 13 FEB 1922 (details about the birth in level 1 above)
2 PLAC Monroeville, Cass Co, MI (more details about the birth)
[This record is about Joseph PRYOR until the next level 0 is reached,
signalling the start of a new record:]
0 NAME Martha /WHITE/ (new record about another individual)
1 SEX F (data about Martha)
1 BIRT (etc.)
```

After the level number, these lines also contain a **Tag**, which is an abbreviation of the type of data in that line. Most tags are obvious; HUSB for husband, PLAC for place, MARR for marriage, etc., but some are more unique, like EMIG for Emigration, and HMOT for Husband of Mother. These tags can also be **Pointers** (@S43@), which indicate another individual, family, or source *within the same GEDCOM file*.

In Reunion, you assign a GEDCOM Tag to fields when you create them. You can also modify these Tags while in the process of exporting a GEDCOM file. However, be aware that not all genealogy programs will recognize your chosen tags, especially if you create one of your own. This could cause the data listed under this tag to be ignored or lost.

When importing a GEDCOM file, a genealogy program uses these level numbers and tags to assemble the data into a family, with relationships intact. The software reads the line numbers and the tags and tries to place the data into the correct fields. If the software doesn't recognize a tag, it either ignores that line or places it in a specific field from which you can move it later. When importing a GEDCOM file to Reunion, you have a choice of either discarding this unrecognized data or placing it into a Note field.

Here is an example of a GEDCOM file, taken from my own family file in Reunion. The **DEST**ination is Ancestral File, and the **CHAR**acter set is MS-DOS.

This first record (a section of the overall file) is the **HEAD**er record (starting with line 0 and going until the next line 0 is detected). The Header contains introductory information, such as the **SOUR**ce program used to create the GEDCOM file, and which **VERS**ion of the program and of GEDCOM was used, etc.:

```
0 HEAD
1 SOUR REUNION (the creating software)
2 VERS V5.0 (the version of Reunion)
2 CORP Leister Productions
```

```

1 DEST ANSTFILE (the receiving/importing software)
1 DATE 12 NOV 1997 (the creation date of the GEDCOM file)
1 FILE Paternal Family File 4.0 (the name of the file on my hard disk)
1 GEDC
2 VERS 4.0 (the version of GEDCOM that was used)
1 CHAR IBM DOS (the character set of the receiving software)

```

The next section is a record for an **IND**ividual, Michael Fitzgerald. He is Individual number **302**.

```

0 @I302@ INDI
1 NAME Michael /FITZGERALD/
1 SEX M
1 BIRT
2 DATE 1842
2 PLAC Duncannon, Ireland
1 DEAT
2 DATE 3 JAN 1916
1 OCCU Theater Stage Hand
1 NOTE Emigrated from Ireland in 1860, prob. landed at the Battery, New York City.
Had a brother who emigrated later and settled in Boston. Michael worked for the old
Forepaugh's Theatre as a stage hand and collected autographs and news clippings of
2 CONT stage stars; spoke with a thick Irish brogue and known for his wit.
1 SOUR @S2@
1 HEAL Died of le grippe.
1 FAMS @F107@

```

[The **CONT**inue tag indicates the the **NOTE** field has more text than will fit under one tag, so it is continued on the next level, 2]

[The **SOUR**ce line shows where the information directly above it was found; this is pointing to **Source number 2**; see below]

[The **FAMS @F107@** is a pointer to a **FAM**ily record; the S on FAMS indicates that the family record is for this individual and his/her **Spouse**, who make up **Family number 107**; see below]

Later in the GEDCOM file, the **SOUR**ces are listed numerically. The following is the record for **Source number 2**:

```

0 @S2@ SOUR
1 NAME "The History of the Fitzgerald Family", compiled by Eileen DeHope and
Elizabeth DeHope Boyce, 1978

```

Still later in the GEDCOM file, the **FAM**ilies are listed numerically. This **Family**, number **107**, is made up of a **HUSB**and, whose **Individual** number is **302**, and a **WIFE**, whose **Individual** number is **303**, and their **CHILD**ren, whose **Individual** numbers are listed. Also included is the couple's **MARR**iage **DATE**:

```

0 @F107@ FAM
1 HUSB @I302@
1 WIFE @I303@
1 CHIL @I38@
1 CHIL @I316@
1 CHIL @I111@
1 CHIL @I109@
1 CHIL @I318@
1 CHIL @I36@
1 CHIL @I321@
1 MARR

```

2 DATE 17 DEC 1862

Remember that in a complete GEDCOM file, many records are included. All of the lines would be listed one after the other, usually with individuals listed first, then sources, and then families. The final record that is required in a GEDCOM file is called a Trailer, and notifies the software that this is the end of the file:

0 TRLR

[Articles Page](#) / [Reunion Home Page](#)

e-mail: info@LeisterPro.com

© 2000 Leister Productions Inc. All rights reserved.