Gedcom

The word **Gedcom** designates a genealogical data exchange format.

It was originally developed by the Mormon Church for religious reasons, then retrieved by genealogists to exchange genealogical data between people running their genealogy applications on incompatible computer systems.

So Gedcom is a kind of genealogical language. It is the most famous and used worldwide.

> The word **Gedcom**, which is an acronym for **Ge**nealogical **d**ata **com**munication, is spelled like an acronym: GEDCOM. By metonymic derivation, a Gedcom also designates a genealogy file in Gedcom format. The file you are working on within Ancestris is a Gedcom.

Since the mid-1990s, with the advent of the Internet and the proliferation of digital exchanges, **the Gedcom specification has gradually become an essential standard for most genealogy software and sites worldwide.**

However, while most of them can export in Gedcom format, **some of them do not strictly respect the Gedcom format** and make some adaptations by adding proprietary structures or using existing ones for different meanings. In some cases, proprietary data structures cannot be properly converted to the Gedcom format and some data might simply not be exported.

> **Ancestris is fully Gedcom compatible with Gedcom versions 5.5, 5.5.1 and 7.0**. As a user, you may safely rely on Ancestris to keep comprehensive genealogy files, without risk of data loss, and share or transmit them with anyone.

> **Version 12 of Ancestris converts your Gedcom file to the 7.0 standard and allows you to manage your genealogy in the 7.0 format**

Characteristics of a Gedcom file

A Gedcom file is a text file, i.e. a file containing human readable lines of text, that can be opened and edited using any text editor, such as Notepad, Kate, Kwrite, Gedit, etc. Its extension name is "***.ged**".

As a result, such a file can be used ***as is*** by any genealogy software, installed under any operating system, without any conversion need.

Each line of text starts with a number and a label. The label is called a "tag". This tag is made up of three or four capital letters. It defines the type of information that follows on the line.

- For example, the tag **PLAC** (= place) always indicates that the text that follows this tag is a place (such as place of birth, place of death, place of a ceremony, etc.)

Records of a Gedcom file

A Gedcom file contains a set of records. A record is a group of text lines, the first one of which starts with a zero "0". A record defines something in particular, which depends on the type of record.

The first and the last record of a Gedcom file are of particular types:

- The first record is called the header (**HEAD** tag) and defines some general information about the file.
- The last record is called the end of file trailer marker (**TRLR** tag). It defines the end of the file.

Each of the other records defines a genealogical entity, with its own set of tags.

A Gedcom file uses 7 entity categories. The records that can be found in a Gedcom file are therefore as follows:

- Records defining **individuals** (tag INDI)
- Records defining **families** (tag FAM)
- Records defining **notes** (tag NOTE or SNOTE)
- Records defining **sources** (tag SOUR)
- Records defining **repositories** (tag REPO)
- Records defining **media files** (tag OBJE)
- Records defining **information submitters** (tag SUBM)

The choice to consider these 7 categories of data as being records is arbitrary of course, but it is always the case when creating a standard.

One could easily imagine other types of records, such as places for example. The fact that a place is not a separate entity does not prevent Ancestris from managing them and respecting the Gedcom format at the same time.

Tree structure of a record

Each record is presented in a tree structure: each tag can include any number of sub-tags.

Sub-tags are hierarchically depending on the previous higher level tag, and may in turn include one or more sub-tags, etc.

Hierarchical levels

Hierarchical levels are numbered.

As each line must imperatively remain in its place from the point of view of the hierarchy, each of them is assigned a number corresponding to the level occupied in the tree structure of the record.

This is why the main level line of each record is level zero, numbered 0. A line located at the level immediately below bears the number 1. A line located at the level immediately below level 1 bears the number 2. And so on.

Identifier and entity records

As mentioned above, apart from **HEAD** and **TRLR** records, all the other records are entity records.

Each entity record starts with a level 0 line followed by the following:

- The ID number of the entity surrounded by two at-signs (@),
- The tag associated with the category to which the entity belongs.
  - For example, the line "**0 @I5@ INDI**" is the first record line of an INDIvidual entity which ID is 'I5'.

Indentation

For greater clarity, lines of a record can be indented so as to more clearly show the relationship between the lines of the record. The information lines underneath a tag qualify the tag.

- Non indented record:

```
0 @I5@ INDI □□□□□=> this defines indidual number 'I5'
1 NAME John Doe □□□□=> The indivudual's name is John Doe
1 SEX M□□□□□□□=> This individual is a male
1 BIRT□□□□□□□=> What follows defines his birth event
2 DATE April 16, 1951□□□=> John Doe was therefore born on April 16, 1951
1 FAMC @F1328@□□□□□=> Family F1328 is the record that defines John Doe's family (FAM) where he
child (C)
```

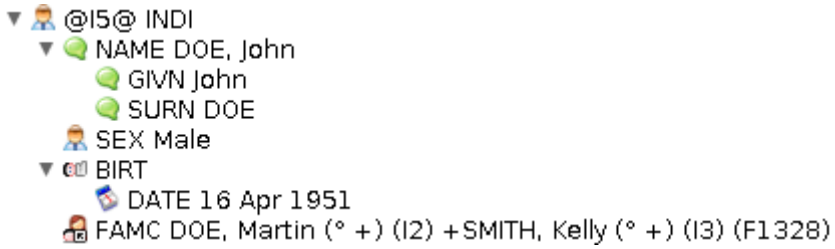
- Indented record:

```
0 @I5@ INDI □□□□□=> this defines indidual number 'I5'
□1 NAME John Doe □□□=> The indivudual's name is John Doe
□1 SEX M□□□□□□=> This individual is a male
□1 BIRT□□□□□□=> What follows defines his birth event
□□2 DATE April 16, 1951□=> John Doe was therefore born on April 16, 1951
□1 FAMC @F1328@□□□□=> Family F1328 is the record that defines John Doe's family (FAM) where he
child (C)
```

The **Gedcom editor** is an editor within Ancestris which shows you the exact information located in the Gedcom file and all of it. It also enhances the display of this information to make it even easier to read.

This editor uses an indented display and does not show level numbers. It also adds handles, to show or hide sub-tag levels, making it easy to expand or collapse any branch.

- This is how the same individual would show in the Gedcom editor:



As you can see, the Gedcom editor enhances the display by adding icons and by fetching relevant hints.

In particular, the "@F1328@" piece of data is replaced, only in the display, not in the real Gedcom file, with the relevant information about the family. Here, we therefore immediately know that John's parents are named Martin and Kelly.

Also the name is split into its lastname and firstname parts.

Line formatting inside a record

Formatting

Each line in a record is made of 3 elements:

- The level number, from 0 to n,
- The **tag** which defines the type of genealogical information that follows,
- The genealogical information.

Example: the line `2 DATE April 16, 1951` means level 2, information of type *DATE* which value is *16 april 1951.*

In order to know what this date refers to, you would have to read the previous lines. Knowing this is a level 2 line, there must be a level 1 (the event in this case) and a level 0 (the record entity) above.

Referencing another entity

Sometimes in a line, one may need to refer to another record entity. This is done using by indicating the identification number of the entity, surrounded by two at-signs (@).

The difference between reference "@id@" defining a record and reference "@id@" pointing to a record is the following:

- If the reference is in the 0-level record before the tag, right after the "0", then it is a definition of a record. Example :
  - `0 @I5@ INDI` : defines individual I5
- If the reference is on the right hand side of the tag, it references the entity. We also say it "points" towards the entity. Example, if we have this line inside an individual record:
  - `1 FAMC @F1328@` : this indicates that the family from which this person is a child (meaning of FAMC), is F1328. It also assumes that F1328 is defined somewhere else in the Gedcom file as a record which should start with `0 @F1328@ FAM`.

Gedcom standard

The Gedcom standard refers to the set of rules that govern what can and cannot be done when writing a Gedcom file, so that everyone organizes genealogical information in the same, so that others can understand. It is therefore the grammar of the Gedcom language.

Three main standards exist, 5.5, 5.5.1 and 7.0, all sharing the same fundamental entities and tags, with a gradual evolution towards better possibilities and clarification. These differences are limited however we recommend the use of Gedcom 7.0 which is more complete and clearer.

> Ancestris handles all standards indifferently and can convert your Gedcom file from one standard to the other and vice-versa.

Gedcom 5.5

> This norm was released on January 2nd, 1996.

You will find details of the comprehensive **Gedcom standard release 5.5 here** in the form of an html site.

Gedcom 5.5.1

> This norm was first released in 1999 as a draft and remained stable until November 15th, 2019 when it was made official.

You can also consult the Gedcom standard release 5.5.1 as a pdf file: **Norme Gedcom 5.5.1**.

These 20 years of stability makes this standard a very well documented and widely used way to exchange genealogical information.

You will find in the 5.5.1 document a comparison between the two Gedcom standards.

Gedcom 7.0.18

> This norm was first released in 2021 and is still evolving within the 7.0.x range.

The specifications of this norm can be found in **The FamilySearch GEDCOM Specification** site.

Revision #21
Created 24 May 2020 16:58:33 by frederic
Updated 26 July 2026 12:10:12 by frederic