

GEDZIP specification version 0.2

22 Feb 2021

Copyright © 2021 The GEDCOM Project.
All Rights Reserved.

This is an early GEDZIP specification.
Download the latest version from The GEDCOM Project,
www.gedcom.org.

table of contents

Contents

GEDZIP specification version 0.2	1
table of contents	3
motivation	4
GEDCOM issue addressed	4
GEDCOM 5.5	4
syncing	4
conceptual solution:	5
GEDZIP	5
use cases	6
GEDZIP basics	7
bundling	7
consistency check	7
GEDCOM version dependence	7
GEDCOM version check	7
file extension	7
directory structure	8
principle of least surprise	8
copying to another system	8
examining third party genealogy data	8
GEDZIP file paths	8
user interface	9
principle of least surprise	9
references	10
GEDCOM specification	10
ZIP format	10
release history	11

motivation

GEDCOM issue addressed

The GEDZIP specification addresses a GEDCOM limitation.

GEDCOM 5.3 (1993) added support for multimedia by allowing references to multimedia files, but no way to transfer multimedia files together.

GEDZIP provides a bundling mechanism; a way to bundle the GEDCOM file and the multimedia objects it references as a single package.

GEDCOM 5.5

GEDCOM 5.3 is a draft. The first GEDCOM specification to allow multimedia object was GEDCOM 5.5 (1996), and it did suffer the aforementioned GEDCOM 5.3 limitation.

The GEDCOM 5.5 specification allows inclusion of multimedia object in GEDCOM files.

As GEDCOM files are text files, the inclusion mechanism requires binary-to-text encoding and decoding. The specified encoding is similar to Base64 encoding.

A practical drawback of all binary-to-text encodings is that such encodings have significant overhead; the size of the text encoding is easily $\frac{1}{3}$ larger than the multimedia file.

A practical issue with the GEDCOM 5.5 binary-to-text encoding and decoding is that very few vendors ever implemented this GEDCOM 5.5 feature.

The GEDCOM 5.5.1 specification (1991), removed the ability include multimedia objects in a GEDCOM file. The "Modifications in Version 5.5.1" section notes the change, without explanation. While GEDCOM 5.5.1 is merely a revision of the GEDCOM 5.5 specification, it still offers several significant new features, so vendors switched to using GEDCOM 5.5.1.

syncing

The GEDZIP specification is twenty years overdue, and may even seem a solution that is not needed anymore. Several major genealogy application allow syncing your genealogy between your desktop application and web app, including the multimedia objects.

However, only a few desktop applications allow syncing, with only a few websites. Syncing technology is fairly complex and its usage in genealogy applications relatively new, so users experience a variety of problems of varying severity. Another complaint is that syncing, even an initial upload or download, is slow.

Perhaps most significantly, while syncing between desktop and web represents a major use case, it is not the only one. Genealogists exchanging information with each other frequently send GEDCOM files, and may like to include referenced multimedia objects automatically.

conceptual solution:

GEDZIP

This GEDZIP specification introduces a simple bundling mechanism; the GEDCOM file and the multimedia files it references are bundled together in a ZIP file.

obvious choice

This bundling mechanism is an obvious choice. The ZIP file format is so ubiquitous that the Windows explorer will let you look inside ZIP files, and programming libraries for zipping and unzipping are readily available.

ZIP files are approachable; they are simple to use and well understood by end-users.

In day to day practice, many genealogist are already using ZIP file to bundle their GEDCOM with multimedia objects when exchanging data with other genealogists.

If a user of a genealogy program that does not support GEDZIP receives a GEDZIP file, they will still be able to use that GEDZIP file.

why bother

The GEDZIP idea is so obvious and so simple that it is reasonable to wonder why we should bother with a specification. The simple answer to that question is that the presence of a specification will encourage vendors to implement the GEDZIP feature.

USE cases

Uses cases for GEDZIP include but are not limited to

- transfer of genealogy data between genealogy applications
- transfer of genealogy data between computer systems
- upload of genealogy data from desktop to web
- download of genealogy data from web to desktop
- initial upload of genealogy data for a syncing system
- initial download of genealogy data for a syncing system
- exchange of genealogy data between researchers

GEDZIP basics

bundling

A GEDCOM file bundles a GEDCOM file and the multimedia objects it references.

A GEDCOM ZIP file should

- include exactly one GEDCOM file
- include *all* the multimedia objects references buy that GEDCOM file
- *not* include unreferenced multimedia objects

consistency check

Products supporting GEDZIP should perform a consistency check before importing a GEDZIP file.

The check to be performed include

- check that there is exactly one GEDCOM file
- check that all referenced multimedia files are included in the GEDZIP file
- check that the GEDCOM zip does not include unreferenced multimedia files

GEDCOM version dependence

The GEDZIP specification does not directly depend on any particular GEDCOM version.

However, GEDZIP can only be used with GEDCOM versions that support multimedia objects, so that is GEDCOM 5.3 and later - but GEDCOM 5.3 is a draft and should not be used.

The first GEDCOM specification to support multimedia objects is GEDCOM 5.5, but GEDCOM 5.5 features a mechanism for embedding multimedia files, so that is the bundling mechanism that should be used with GEDCOM 5.5.

Products supporting GEDZIP should create GEDCOM files according to GEDCOM 5.5.1 or later. As of 2021, GEDCOM 5.5.5 is the preferred GEDCOM specification.

GEDCOM version check

Products supporting GEDZIP should perform a GEDCOM version check before importing a GEDZIP file, and issue a warning or error if the GEDCOM version is less than 5.5.1.

Products supporting GEDZIP may demand GEDCOM 5.5.1 or later; the product need not support GEDCOM 5.5 or earlier in GEDZIP files.

file extension

GEDZIP file do not have their own file extension, but simply use the standard .ZIP extension. This ensures that users understand that they can unzip the file.

GEDZIP files use the same base name as GEDCOM files; supporting applications simply create *filename*.ZIP instead of *filename*.GED.

directory structure

principle of least surprise

GEDZIP follows the principle of least surprise; i.e. the specification tries to ensure that things work the way an average user expects.

copying to another system

A user who uses GEDZIP to transfer genealogy data from one computer to another expects their directory structure to be preserved.

However, a specification such as GEDZIP cannot guarantee exact duplication of a directory structure, as different computers may not support the same drive letters.

examining third party genealogy data

A user receiving genealogy data from another researcher expects to be able to unpack and then view that data, and to be able to easily delete it all when they are done with it.

A user choosing to use a third party GEDZIP file does not want third party files copied all over their system, and certainly does not want many new top level directories created for it. A user expects all the GEDZIP files to be contained in one directory (and subdirectories), for easy deletion.

GEDZIP file paths

- The GEDCOM file should not merely contain filenames, but file paths
- These file paths should be relative (to a top directory)

Applications that import GEDZIP files should allow the user to choose a project name and then unzip the GEDZIP file in a directory for that project.

Applications should unzip all GEDZIP files to a directory the same way a ZIP utility would do so, had the user had chosen to unzip the GEDZIP file using such a utility.

Applications bundling multimedia objects keep this in mind; they try to maintain the original file paths as much as possible, while making sure all file paths are relative to a top directory.

user interface

principle of least surprise

The user interface for supporting GEDZIP should follow the principle of least surprise; the user interface for importing or exporting GEDZIP files should be similar to the user interface for importing and exporting GEDCOM files.

It is recommended that the user interface for exporting GEDCOM or GEDZIP files be one and the same, with an option that lets the user choose between GEDCOM and GEDZIP export with labels such as "GEDCOM (no multimedia)" and "GEDZIP (multimedia included)".

Both GEDCOM import and GEDZIP import should prompt the user for a project name, with that name used for the project's directory.

references

GEDCOM specification

The current GEDCOM specifications can be found on *The GEDCOM Project* website, www.gedcom.org.

ZIP format

The ZIP format is defined in ISO/IEC 21320-1:2015.
<https://www.iso.org/standard/60101.html>

release history

date	version	brief description
2021-02-21	0.1	Initial draft
2021-02-22	0.2	Added references section and release history