Wesley Johnston's Guide to Getting Started with Graphs for Genealogists (2023 Edition)

Wesley Johnston, last updated 7 May 2023

The following are the steps that I use to install Neo4j and Graphs for Genealogists and prepare and upload my data to the database. These instructions are for GFG 1.3.0 or later and for Neo4j 5.7.0 but not later. Neo4j setup has a different numbering so that you will use Neo4j Desktop Setup 1.5.8 or later.

I have also added instructions in the appendices for two scenarios. In one scenario, I want to upload a clean new copy of my data but do not want to reinstall Neo4j nor GFG. In the other scenario, I want to completely remove and install Neo4j and GFG with the latest versions. These must be done properly since they can lead to corrupting your environment if not done properly. So, if you already have an installation of Neo4j and GFG and are in one of those scenarios, see the appropriate appendix instead of using the following initial installation and setup instructions.

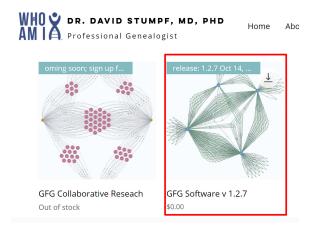
As new versions of Neo4j and GFG come out, these instructions may need updating.

A separate guide is needed for what to do next, dealing with all the generated spreadsheets, adding a "grass" file so that everyone is looking at the same styles when collaborating and using the GFG reports and working with your database in the Neo4j browser. But that does not exist yet.

IMPORTANT CONCEPT: GFG is an application that runs on top of Neo4j. So, GFG cannot work unless Neo4j is up and running.

Step 1: Download the latest version of Graphs for Genealogists.

The GFG shop holds the latest version of GFG. Go to https://www.wai.md/shop and click on the icon for the GFG software. Here is how it looks as of a prior writing. You should see version 1.3.0 or later.



Step 2: Install Graphs for Genealogists.

Unzip the downloaded file. Run the executable file ("Graphs for Genealogists.msi") that you downloaded in step 1. For the moment, there is no need to start GFG. The installation is all that I do in this step. I do not start GFG.

Second edition update: If you are running Norton or other security software, you may need to make C:\Program Files (x86)\WAI\Graphs for Genealogists\GFG_2022.exe a trusted application so that your security software does not block it from accessing your files.

Step 3: Download the Neo4j Software and the Activation Code and Install Neo4j, which will run it.

Download the Neo4j Enterprise edition from https://neo4j.com/download-neo4j-now and be sure to save the very long activation code which will speed up installation of Neo4j. I save the activation code into a text file. Wait for a bit after the installation completes so that Neo4j can check for updates for you to install. (I do stop the example Movie DBMS, but that is not really necessary since it will be stopped when you start your own DBMS.) This process installs the Neo4j server on your PC.

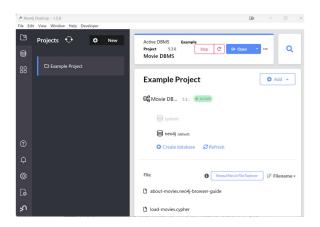
When I am doing a reinstallation of the full environment, I usually do not do a new download from the Neo4j website since the executable file that I had already downloaded works fine for a re-install as long as I still have the activation code from that download. Once I re-install Neo4j, I let it sit for 10 minutes or so, which allows Neo4j to check for updates which I can then install so that I still have the latest versions of the components even though I did not do a new download of the executable file.

In Windows 11, during the Neo4j installation, about when it reaches the creation of the example database, Microsoft pops up this PowerShell window.



This is new either to Windows 11 or to version 1.5.8 of the Neo4j Desktop installer. I simply close this window.

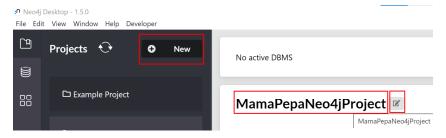
The Neo4j installation ends by running the Neo4j software and opening the Neo4j Desktop with the Example Project's "Movie DBMS".



Step 4: Create and Name Your Neo4j Project.

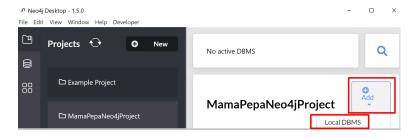
It is important to keep separate the term project in the context of Neo4j and of GFG. Each has a project which requires a name. Your GFG project will have the same name as the database you create in Neo4j. The Neo4j database is in a Neo4j DBMS (database management system) which can have multiple databases. And the DBMS is within a Neo4j project. The Neo4j server can have multiple projects (DBMS), and each DBMS can have multiple databases. The reason this is important is that professional genealogists might want to have separate DBMSs for different projects.

To create my Neo4j project, I click on "New" in the top left "Projects" section which generates a new Neo4j project. I can then click on the editing icon (in the small red box in the image below) to the right of the project name (visible when I hover over the project name) to give my project the name I want.

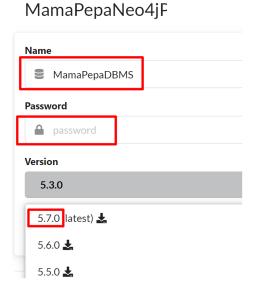


Step 5: Create the Neo4j DBMS for Your Neo4j Project, Add the PlugIns and Start the DBMS.

To the right of your project name in the right side of the Neo4j Desktop, click the "Add" button. This will open a pull down list. Click on "Local DBMS" in the list.



This will alter the Neo4j Desktop to look like this.

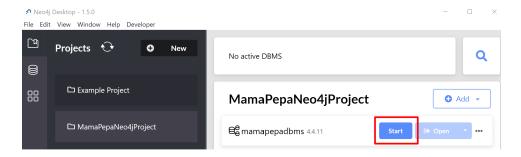


Enter the name you want to give to the DBMS (for example in this case, I called it "MamaPepaDBMS"). Enter a password. I keep it very simple since it really is only used by GFG in the setup and access.¹

Then click the pulldown triangle/arrowhead for the "Version" section and choose the 5.7.0 version.

Once you have set the version, it will be downloaded if not already on your PC, and then the list will close, and you will see the "Create" button to the lower right. Click on the "Create" button.

Once the DBMS is created and both PlugIns installed, I start the DBMS. Hover over the name of the DBMS so that the Start button appears to the right, and click on the Start button.



Step 6: Stop the DBMS, Install the PlugIns and Re-Start the DBMS.

Graphs for Genealogists requires two Neo4j PlugIns. First stop, the DBMS by clicking on "Stop" and waiting for the stopping of the DBMS to complete.

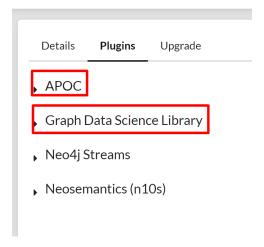


A red button will appear by the DBMS name until it has stopped. You will know for certain that it has stopped by hovering over the name of the DBMS and seeing the "Start" button on the right has re-appeared.

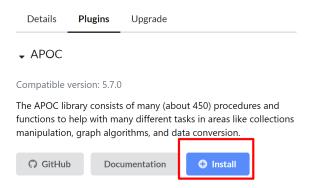
Now click on the DBMS name. This will open a panel to the right (if you do not see it, expand your window to full screen) where you click on the "PlugIns" tab.



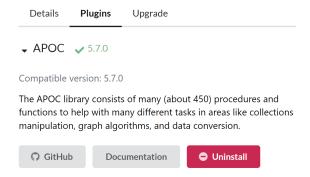
This will open the list of available Neo4j PlugIns. You need to install APOC and Graph Data Science Library.



These each take a while to install. So, start by clicking on "APOC". This will open a new section where you click on "Install".

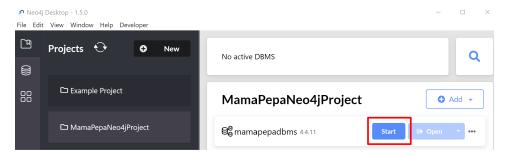


Once it is installed, a green check mark will appear by the name, and the blue Install button will turn to a red Uninstall button.



Now click on "Graph Data Science Library" and install it. Once it is installed, you can click on the X at the top right of the right side pane to close it.

Once both PlugIns are installed, I re-start the DBMS. Hover over the name of the DBMS so that the Start button appears to the right, and click on the Start button.



When I restart the DBMS this time, another Microsoft PowerShell window pops up, and I simply close it.

MILESTONE REACHED: This completes the Neo4j setup other than the creation of your database. Before you do that, you need to switch over to the GFG software to begin connecting GFG to Neo4j.

Step 7: Run the GFG Harmonization.

I wait until a later step to set up the database. Before I do that, I run the GFG Harmonization.

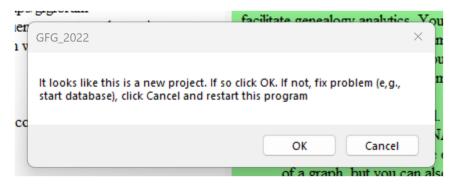
The GFG Harmonization does the following to enable the GFG capabilities:

- 1. Modify the Neo4j conf file, including the memory allocated to Neo4j and how it manages transaction logs.
- 2. Add an apoc.conf file.
- 3. Create a new c://Genealogy folder and a set of subfolders and files used by GFG.

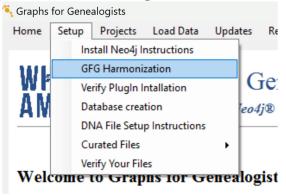
Just to be sure, especially if I have been away from the computer since the prior step, I stop and restart the DBMS in Neo4j. In fact, I will sometimes shut down Neo4j completely and then restart it and start the DBMS just to be sure everything on the Neo4j side is active. I then start GFG.



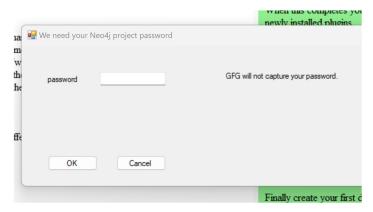
Since I do not yet have a GFG project, I click OK on the initial popup box.



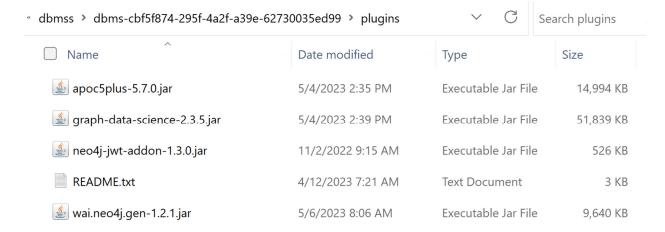
Then I use the Setup pulldown menu at top left and click on "GFG Harmonization".



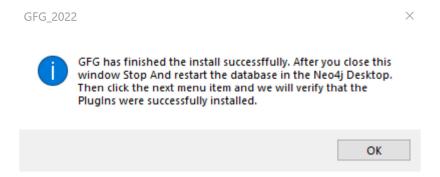
GFG will ask you for your Neo4j DBMS password, since it needs to integrate its GFG files with your Neo4j files.



Once you click on the "GFG Harmonization" option in the setup, it will start working with a popup message that tells you that you can watch the installation in the plugins folder. A window will pop up for the plugins folder of your DBMS. You will see it populate the folder with GFG JAR file (wai.neo4jgen-n.n.n,jar). Here is how mine looks.



Once these files are in the folder, the step is done. Be sure to wait while it runs since it will appear at times to not be doing anything, but it will still be working until you receive this popup window. This can take some minutes.

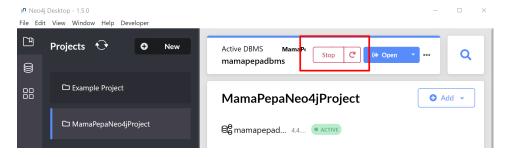


Once it is done, you can close the window that shows the plugins folder.

Step 8: Stop and Restart Your Neo4j DBMS. (Easy step to forget to do.)

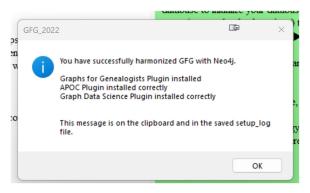
Once the popup window tells me that GFG Harmonization has successfully completed, I return to the Neo4j Desktop to stop and restart my DBMS. When the DBMS restarts, it will be able to use the GFG and other plugins installed by the GFG Harmonization.

In the Neo4j Desktop, click on the "Stop" button.



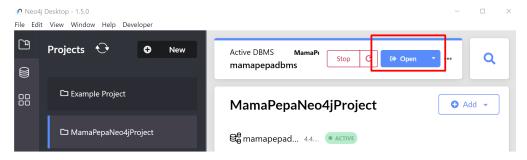
Step 9: Verify Plugin Information.

Back in the GFG app, I click on the "Setup" menu item for "Verify Plugin Installation". Sometimes this comes back saying that the GFG plugin is not installed while the other two are installed. If this happens, I simply click on "Verify Plugin Installation" again, and the new popup shows that everything is properly installed.



Step 10: Create Your Neo4j Database.

To create your Neo4j database, you use the Neo4j browser for your DBMS. You open the Neo4j Browser from the Neo4j Desktop by clicking on the open button once you have started your Neo4j DBMS.¹¹



This opens the Neo4j Browser in a separate window. In the Neo4j Browser (not the Neo4j Desktop), click on the database icon at top left. Then make sure you choose the "system" database and enter "create database xxx" where "xxx" is the name (must be all lower case and only letters) you want for your database. Then click the blue arrowhead (what David Stumpf calls the "Go button").



Be sure to remember this name and the password you used for you Neo4j DBMS since you will need these to set up your GFG project. You can now close the Neo4j Browser (but NOT the Neo4j Desktop) if you want.

MILESTONE REACHED: This completes the Neo4j and GFG setup. Now, you must prepare your data, define your GFG project and then load your data into your database.

Step 11: Prepare Your Data.

Now, you step back from both Neo4j and GFG to do your data preparation (although you will need to obtain the template for the GEDCOM_DNA spreadsheet from GFG). If you have a lot of DNA test takers among those in your family tree, this can be the most time-consuming step. You have to prepare all three types of data:

- 1. Your GEDCOM
- 2. The Family Tree DNA files of your DNA testers
- 3. A curated file to tell connect each test taker's FTDNA files with their record in the GEDCOM

Step 11A: GEDCOM

Download your GEDCOM from your family tree database. Place the GEDCOM in your computer's C:/Genealogy/gfg/gedcom folder.

Step 11B: Family Tree DNA Files

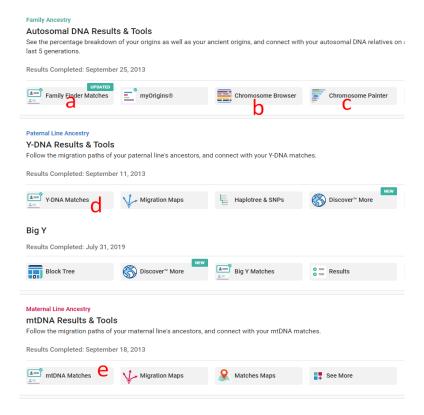
Now you need to create the directory structure for your DNA files. These need to be in their own folder (recommend c://Genealogy/gfg/dna). Within this folder you will create a subfolder to hold each kit's data. You can use whatever name helps you manage these folders, but I use the name of the person, as described below. You can create these folders in advance or at the time you download the DNA data files.

At this time, GFG works only with Family Tree DNA files. For each of your DNA testers whose FTDNA account you can access or who are willing to send you their files, download the following CSV (comma-separated variable) files from the FTDNA account of each person. These are optional. Some kits may only have Y-DNA (file d) or mtDNA (file e). Some will only have files a-c. Some will not have activated a transferred kit and will only have item a available. You can use any combination, but if the tester did do a Family Finder test, you should do files a-c for that user (and either d and/or e if they did one or both of those tests).

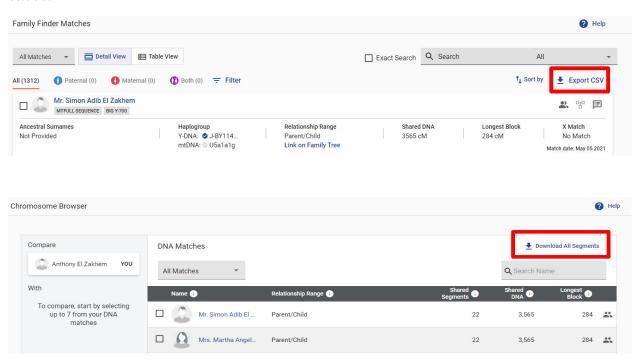
You can use any combination

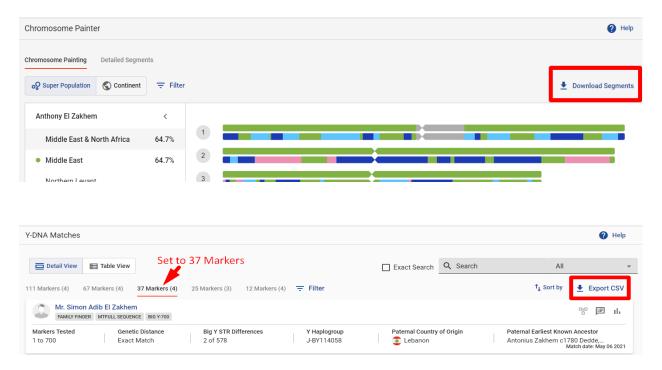
- a. Family Finder Matches
- b. Chromosome Browser
- c. Chromosome Painter
- d. Y-DNA (NOT Big Y) 37-marker Matches
- e. Mitochondrial DNA Matches

The following image shows the location of each type of file on the account's FTDNA dashboard.



The following show the locations where to click to download the file in each of the five cases.





Sub-Folder Names

Each set of files goes into a separate folder in the C:/Genealogy/gfg/dna folder. The folder containing a person's FTDNA files can be any name, but I use the person's name. GFG iterates through the folders and gets the kit number, stopping when it finds the one in the queue. Then it processes the files for that kit.

Step 11C: Curated File

The curated file tells GFG which FTDNA kit goes with which person in the GEDCOM file. It does this by pairing the FTDNA kit number with the DNA test taker's record number (RN) in the GEDCOM file. This requires care in finding the correct RN for the person in the GEDCOM file.

The curated file can also contain the matches who are in the GEDCOM file but whose DNA kit is not loaded (e.g., you cannot access them). The link between the Person and DNA_Match nodes (in the Neo4j database) is then done using the RN and the curated name match. The curated name in such cases is often different from the name in the Person node.

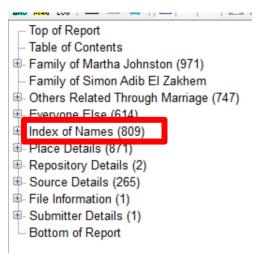
The GEDCOM uses an "@I###@" (capital letter i) followed by a number. For the curated file, you only want the numeric part of the number ("###") and not the "@" or "I". So do not confuse the "I" with a digit "1".

Your family tree software may allow you to see the person's RN (record number). For example, this is possible with RootsMagic and Legacy. However, I keep my master trees online on Ancestry and then use Family Tree Maker to download them, and FTM has no way (according to those on the FTM Facebook group) to show you the record number for a person. So, finding the record number in my FTM-exported GEDCOM file can either be done with a text editor or some third-party software.

I prefer to use Louis Kessler's Behold! Software (www.beholdgenealogy.com) to find the person's RN in the GEDCOM file. Behold! allows you to make sure that you have the correct person when there are people with the same name in the GEDCOM. I will show both the Behold! method and how you can find the RN for a person with a text editor such as Notepad.

Behold! Method

Start Behold! and open your GEDCOM file from Behold! Behold! instantaneously reads your GEDCOM and opens your dashboard. Double-click on the "Index of Names" option in the left-side list.



The double-click opens both the right-side Index of Names that shows every name and the left side list of surnames. I find it easiest to use the left side to go to the surname I want and click on it and then select the person that I want from the right-side list. Clicking on the person in the right-side list opens their individual record where you find their "Person ID" on the line after their name at the top. Remember to use only the numeric portion of the Person ID for the curated file and also not to confuse the capital I before the number for a number 1.



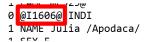
So, in this example for Julia Apodaca, the Peron ID is I1606. You use only the numeric portion (1606) for the curated file. So, the curated file line for Julia Apodaca would look like this.

	_	_	_	_	-
Match_Name	Curated_RN	Kit	File_Upload	Notes	email
Julia Apodaca	1606	123456	Υ		

Enter her name, her RN that you just looked up and her FTDNA kit number. Also be sure to put a "Y" in the File_Upload column since you are uploading her FTDNA files.

Text Editor Method

If you use notepad, you have to search for the name of the person. You are looking for that person's very first line, which has a 0 (zero) in the left column and "INDI" to the right. You want the number on that line that is preceded by the capital letter "I".



So, in this case, you use "1606" as the RN in the curated file, as in the Behold! method.

Use Correct Names

Open one person's Family Finder CSV file with Excel to see the first, middle and last name of each of the person's matches who you are also uploading. Use the names as they appear in those three fields for the person's name in the curated file (see below). Do NOT include any other text that appears in the Full Name filed such as Mr. or Mrs. even if it appears in the match's full name field.



GFG matches the name in the curation file to data in the DNA files. It's very important they are exactly the same. Processing errors will occur if there are double spaces or if a middle initial period is in the DNA match list but not in the curated file or vice versa.

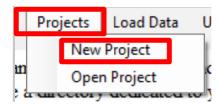
Proper Saving, Naming and Placement of the Curated File

While you can use Excel to edit the curated file, you must save it as a CSV (comma-separated variable) file for the upload. If you want to stick with the default file name (which is what I do) that will come up when you define your GFG project, you must rename it as "XXX_curated_data.csv" where XXX is the name of the database (not the DBMS) that you created in Neo4j in step 10 above. The file can have any name, as long as it has no spaces, but if you do not use the default name, then you must find that file in the GFG project definition step.

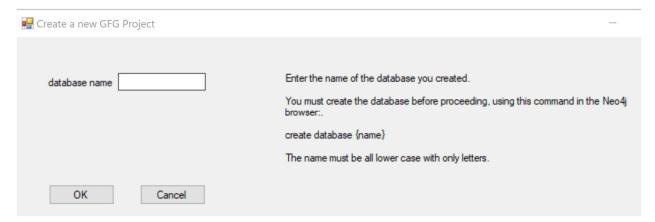
Finally, place this file in your C:/Genealogy/gfg/curated folder.

Step 12: Create Your GFG Project and Find the Directory Path for your GEDCOM file

On the GFG top menu, click on "Projects" and then on "New Project".

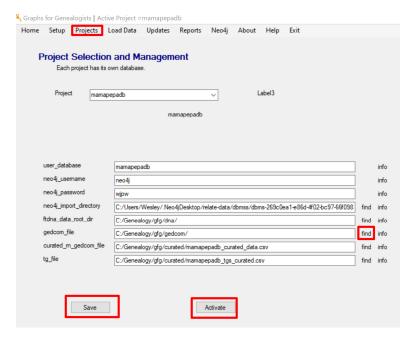


This opens the start of the creation of the new GFG project with this window.



Enter the exact same name that you used in step 10 above when you created your Neo4j database (not the Neo4j DBMS) with the lower-case name.

GFG then takes you to the project definition page. If the project name is blank, use the pulldown arrow to choose your project name.



All of the default settings are complete since you have correctly named and placed the files in the prior steps. (If you used other names, you will have to navigate to find the file and select the path or file.) There is one exception because while GFG knows that you GEDCOM file is in the C:/Genealgoy/gfg/gedcom folder, it does not know the name of the file. So, use the Find button to find your GEDCOM file so that this line of the project settings includes your GEDCOM file.

It is important to note that none of your changes to the settings are automatically saved. You must click the "Save" button to save them. Once you have all of the settings the way you want them, click on the "Activate" button. Both the "Save" and the "Activate" clicks open a window with the settings of the project in a text file. Both

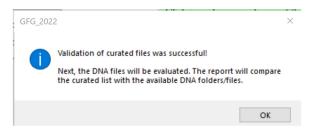
the "Save" and "Activate" buttons may generate warnings for things like a file name with spaces; follow the directions to address these messages.

Step 13: Verify Your Files.

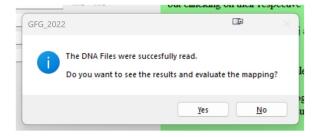
In the GFG "Setup" menu, click on "Verify Your Files". This will check that your files are properly named and in the correct folders, as defined in the project settings. You can ignore the message about not finding a triangulation group curated file when you see this popup that the file is optional and can be ignored.



Click OK, and you will see the following popup to tell you the files are verified.



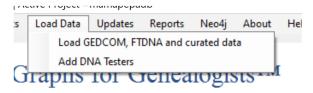
You will again be asked for your password, and this window will pop up.



If there is some row that does not map into a "Folder_Name", then you need to either correct the name in the curated file or else change the name of that person's folder in the FTDNA file folders.

Step 14: Load Data.

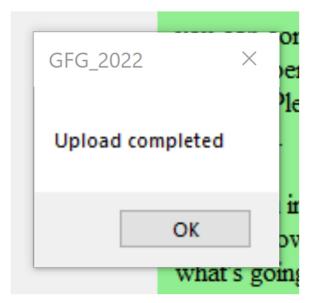
At last, you are ready to load your data. Click on the GFG "Load Data" menu item and then on "Load GEDCOM, FTDNA and curated data".



Be sure to enter the ancestor record number. This person will seed the initial creation of Avatars. So, make sure he or she is one of your research subjects. You will be able to change this later without otherwise changing things. I choose an ancestor several generations back on one of the ancestral lines.



The load process can take about 10 minutes. It will show the import folder being populated. At times, nothing will change in that folder. It will also open an Excel spreadsheet with several tabs and then several more spreadsheets. Wait until you receive the message that the load has completed.



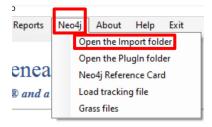
Appendix 1: Reloading Your Database after Data Correction

When you want to keep the same installation environment (Neo4j and GFG) and want to upload a fresh version of your database, follow the steps in this section.

Step 1: Delete the contents of the Import folder for your database.

You really do not have to do this step, but it makes the import folder cleaner if you clear it out before dropping and re-creating your database. You can either delete everything in the Import folder or you can create a sub-folder and move all the files from the old version into that folder so that the Import folder is cleared out for the reloading of your data.

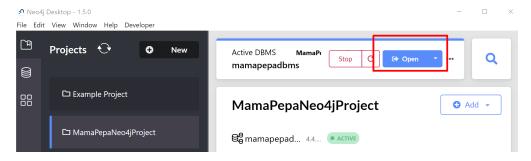
The easiest way to open the Import folder is in GFG. Click on "Neo4j" and then on "Open the Import folder".



Do NOT delete the Import folder itself.

Step 2: Drop your database.

To create your Neo4j database, you use the Neo4j browser for your DBMS. You open the Neo4j Browser from the Neo4j Desktop by clicking on the open button once you have started your Neo4j DBMS.



This opens the Neo4j Browser in a separate window. In the Neo4j Browser, click on the database icon at top left. Then make sure you choose the "system" database and enter "drop database xxx" where "xxx" is the name you had used for your database. Then click the blue arrowhead.

Step 3: Create your database again

In the same place where you just dropped your database, enter "create database xxx" where "xxx" is the name (must be all lower case and only letters) you want for your database. You can either edit the just-entered drop command and change "drop" to

"create", or you can enter the entire command into the blank entry box at the top of the stack.

Step 4: Reload your database.

Repeat what you originally did in step 14 in the main section.

Appendix 2: Starting Completely Over with Neo4j and GFG Installation

If you want to completely clear your environment and start again from scratch, use these instructions.ⁱⁱⁱ

Step 1: Uninstall Neo4j and Graphs for Genealogists.

This is the expected step. When you uninstall Neo4j, it will ask you a question about other files. I always click yes to remove everything that the uninstall process can remove.

Step 2: Delete the Neo4j Desktop Folder and the Genealogy Folder.

This is the UNexpected step. You must delete these two folders to go back to a clean environment.

The ".Neo4jDesktop" folder is in the User folder with for you on this PC. So, it is in C:/Users/XXX where XXX is your name on your computer. Simply delete the ".Neo4jDesktop" folder in its entirety.

GFG creates the C:/Genealogy folder. Simply delete the C:/Genealogy folder.

Step 3: Start at Step 1 of the initial installation and continue from there.

You have now restored your environment to how it was before you first installed Neo4j and GFG. So, go to step 1 of the main instructions above to do the new installation and setup.

¹ GFG creator David Stumpf adds "A bit more on passwords. Each DBMS can have its own password. Each database within a DBMS will have the same password. GFG uses this password every time it accesses the database. It's an important security feature. Although the risk is low on a PC, it's not correct to say GFG does not use the password."

ii GFG creator David Stump adds that there is a second option for opening the Neo4j browser: "from your browser (Firefox, Chrome), navigate to http://localhost:7474/browser/. The first time you do this, use the password "neo4j". Personally, I prefer [this method] for most of my work." iii I am including this here because from time to time, I need to be able to see my environment as a GFG newbie sees it. The only time that I used it other than for helping newbies was when I

thought that I had corrupted my installation and wanted a complete fresh clean start. As GFG creator David Stumpf points out: "This should NOT be something they should do. You can reinstall Neo4j without deleting things. The advantage of this is that you preserve all your prior work setting up GFG projects. The only thing you may have to change is the file with the database path."