



*Tom Nicolai & Kati Mozygamba*

**User Manual**  
**QualiAnon – Tool for the anonymization of text data**  
**(v1.3)**

University of Bremen – Qualiservice Technical Report  
**QS-TR-2-2023**

Empfohlene Zitierung / Suggested Citation:

Nicolai, Tom & Mozygemba, Kati (2023): QualiAnon User Manual, v1.3 Qualiservice Technical Report 2-2023, Bremen, doi: <https://doi.org/10.26092/elib/2576>

Software available at: <https://github.com/pangaea-data-publisher/qualianon>

Wir entwickeln die Software permanent weiter, so dass Sie ggf. aktuellere Einträge im in folgendem Support-Dokument (living document) finden:

[https://docs.google.com/document/d/1fLLYvsgXjh\\_p9p\\_E1fhikkIPQb19VUiltbRgaWQoD-M/edit#heading=h.3vpz66t09gfa](https://docs.google.com/document/d/1fLLYvsgXjh_p9p_E1fhikkIPQb19VUiltbRgaWQoD-M/edit#heading=h.3vpz66t09gfa)

© FDZ Qualiservice, Oktober 2023

### **Impressum**

Forschungsdatenzentrum (FDZ) Qualiservice

Universität Bremen

SOCIUM – Forschungszentrum Ungleichheit und Sozialpolitik

Mary-Somerville-Str. 7

UNICOM - Gebäude

D-28359 Bremen

Germany



Webseite: <https://www.qualiservice.org>

E-Mail: [info@qualiservice.org](mailto:info@qualiservice.org)

# Inhalt

<b>1</b>	<b>Introduction .....</b>	<b>5</b>
1.1	Getting started.....	7
1.2	Release notes .....	7
1.3	Citation.....	7
<b>2</b>	<b>Welcome screen .....</b>	<b>7</b>
2.1	New project .....	8
2.2	Open project.....	8
<b>3</b>	<b>Main project window .....</b>	<b>8</b>
3.1	Window menu .....	9
3.1.1	File .....	9
3.1.2	Edit.....	9
3.1.3	Window .....	9
3.1.4	Help .....	9
3.2	Project structure tabs .....	9
3.3	Document area.....	10
3.4	Bottom tab area.....	10
3.5	Status line .....	10
<b>4</b>	<b>Create and edit categories.....</b>	<b>10</b>
4.1	Create a new category .....	11
4.2	Organize categories .....	11
4.3	Create labels for a category .....	12
4.4	Edit labels of a category .....	12
4.5	Edit the choices of a label.....	13
4.6	Load a list for label .....	14
4.7	Load a list for label lookup.....	14
4.8	Using custom lists for replacement.....	17
<b>5</b>	<b>Working with replacements and markers .....</b>	<b>18</b>
5.1	Add a marker with a new replacement.....	19
5.2	Populate labels from a replacement list.....	21
5.3	Mark multiple occurrences of a word or phrase with a new replacement .....	23
5.4	Add markers to an existing replacement.....	24

5.5	Change marker to selection.....	25
5.6	Reassign markers to another replacement.....	26
5.7	Find replacement and label information for a marker.....	28
5.8	Edit a replacement.....	29
5.9	Remove a marker.....	30
5.10	Remove multiple markers.....	31
5.11	Remove replacement with all related markers.....	32
<b>6</b>	<b>Find text.....</b>	<b>33</b>
<b>7</b>	<b>Creating and configuring exports.....</b>	<b>34</b>
<b>8</b>	<b>Project directory and files.....</b>	<b>35</b>
8.1	Directory layout.....	35
8.2	Get the exported documents.....	35
8.3	Reuse categories for another project.....	36
8.4	Archive a project securely.....	36
8.5	Remove the originals, but keep the labels.....	36
<b>9</b>	<b>Bug reports and feature requests.....</b>	<b>36</b>

## 1 Introduction

Research concerned with living persons usually contains sensitive information. Due to this sensitivity the processing of this kind of research data requires to take legal regulations as well as ethical standards into account. Anonymization and pseudonymization are central instruments to fulfill these requirements to protect personal rights of research participants, researchers and third persons. Both instruments are therefore common in empirical research with and about humans - independent of subject and discipline. However, protecting research data is only one side of the coin when you think of data sharing and reuse. Hiding, or deceiving information might prevent research or lead to wrong results. Especially if data are supposed to be shared for scientific re-use, researchers have to deal with two potentially conflicting interests: protecting personal data and enabling further research.

To address both – data protection and scientific re-use – the Research Data Center Qualiservice developed the concept of „Flexible Anonymization“ (Kretzer 2013<sup>1</sup>, Mozygemba & Hollstein 2023<sup>2</sup>), replacing sensitive personal data with more abstract information relevant for further research on different levels of abstraction, which is included in QualiAnon – a tool for the anonymization of text data.

QualiAnon allows for semi-automatic anonymization/ pseudonymization keeping the user in control. It enables to mark and replace sensitive information and allows for different levels of abstraction as well as for replacements on case basis or on study level (i.e. across all cases). QualiAnon provides versions of anonymized datasets on different abstraction levels. The user can establish systems of replacement categories, that can be shared and further developed and thus be the basis for topic related anonymization standards, which allow for a high amount of reusability. The export of pseudonymized data is also possible. The tool allows for the integration and creation of lists to support the replacement process; it allows to share and re-use the developed replacement categories, anonymized data, replacement lists, exports etc. are stored separately.

---

<sup>1</sup> Kretzer, S. (2013). Arbeitspapier zur Konzeptentwicklung der Anonymisierungs-/Pseudonymisierung in Qualiservice. [https://wiso.genios.de/document/SSOA\\_\\_47605](https://wiso.genios.de/document/SSOA__47605)

<sup>2</sup> Mozygemba, Kati & Betina Hollstein (2023): Anonymisierung und Pseudonymisierung qualitativer textbasierter Forschungsdaten – eine Handreichung. Qualiservice Working Papers 5-2023, Bremen, <https://doi.org/10.26092/elib/2525>

The tool works with a system of Universally Unique Identifiers (UUIDs) (see Figure 5). The anonymized interview ("blackened interview") contains the marker ID that is set in the original text during editing. The replacement ID refers to the translation selected in each case. Marker IDs and replacement IDs of a study as well as the original information are stored in another document.

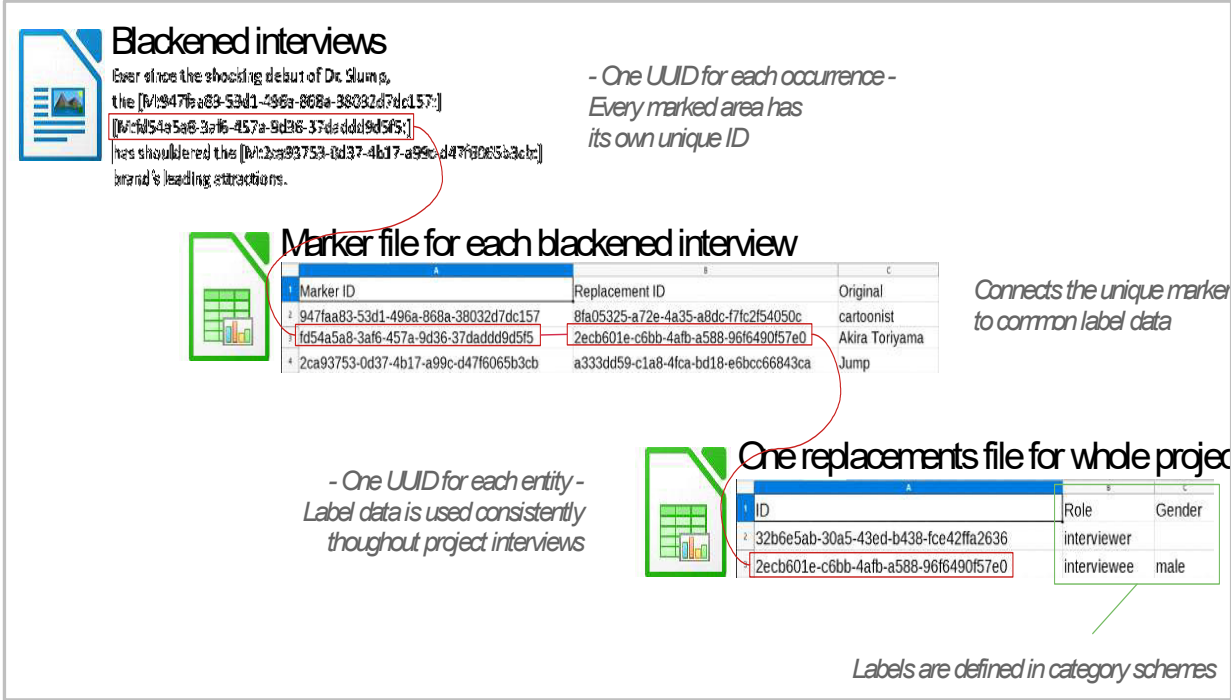


Figure 5: The data model of QualiAnon - connection of UUIDs

## 1.1 Getting started

QualiAnon runs on Windows, MacOS and Linux. It requires Java to run.

1. Go to <https://www.java.com>
2. Download the latest release of Java 8 and install according to the instructions
3. Download the latest release of QualiAnon (quali-anon-v1.2.2.jar) from <https://github.com/pangaea-data-publisher/qualianon/releases>. The repository is not openly accessible. Please drop us a message with your email or github user name and we will grant you access. Contact us at <https://www.qualiservice.org/en/contact.html>.
4. Open quali-anon by double-clicking the jar-file. Alternatively, use the command-line of your OS: "java -jar quali-anon-v1.2.2.jar"

QualiAnon will probably not run on newer or older versions of Java. Also, alternative Java distributions might not work. If QualiAnon does not launch correctly, check if there might be other Java versions installed on your computer and uninstall them.

## 1.2 Release notes

Please see <https://github.com/pangaea-data-publisher/qualianon/releases> for release notes and changelog.

## 1.3 Citation

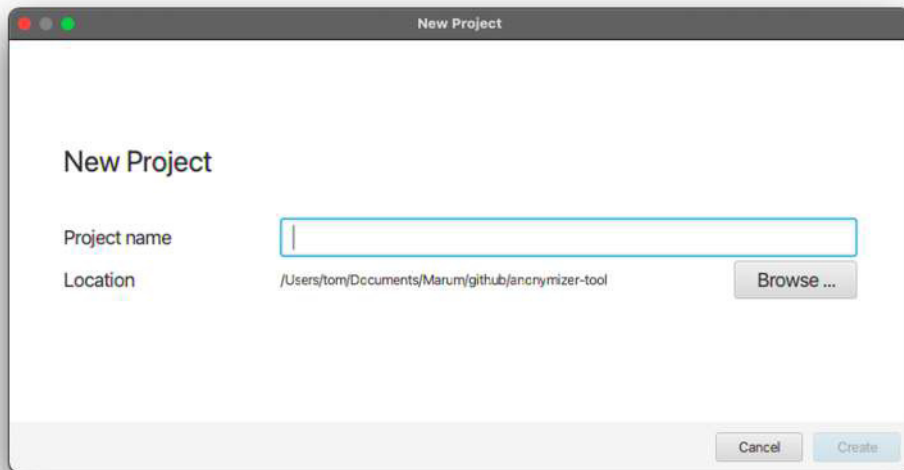
Use the following citation to reference QualiAnon in your works:

Tom Nicolai, Kati Mozygamba, Susanne Kretzer, Betina Hollstein (2021): QualiAnon - Qualiservice tool for anonymizing text data (version 1.1.0). Qualiservice. University of Bremen. Software available at: <https://github.com/pangaea-data-publisher/qualianon>

## 2 Welcome screen



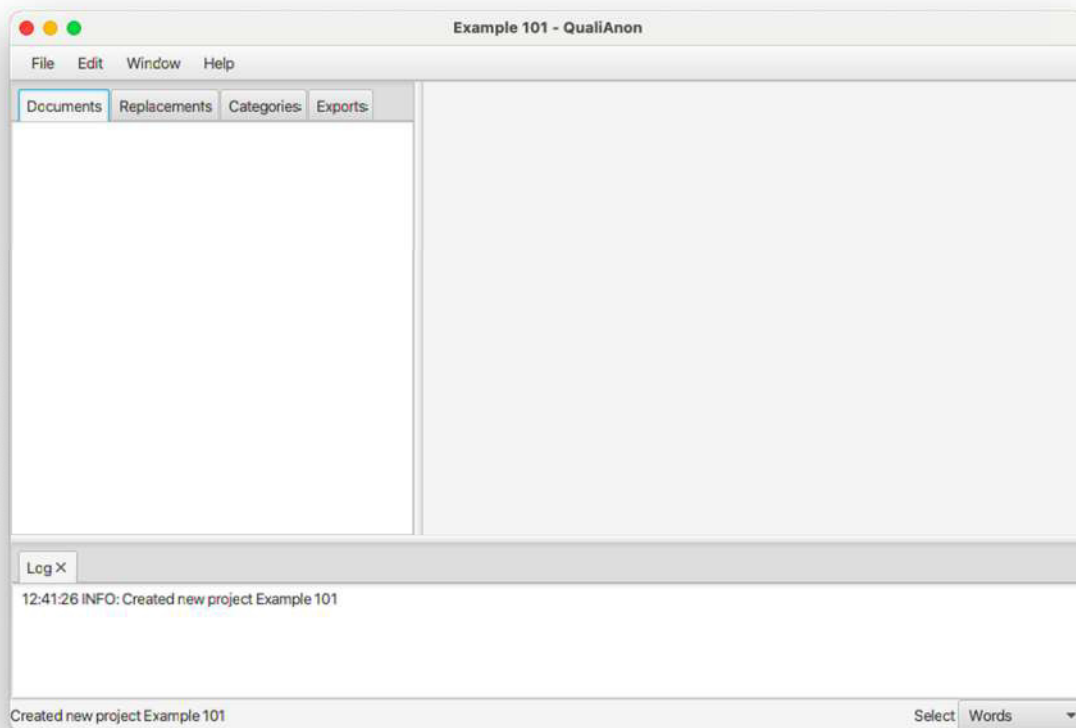
## 2.1 New project



## 2.2 Open project

Select the project folder to open an existing project.

## 3 Main project window





### 3.1 Window menu

#### 3.1.1 File

- **Open Project:** Closes the current project and opens another project.
- **Close Project:** Closes the current project and returns to the welcome screen.
- **Save File:** Saves changes to the currently displayed document.
- **Save All Files:** Saves changes to all documents in the project.
- **Import Document:** Import one or multiple documents into the project. QualiAnon imports documents in the .docx format, which can be created with Microsoft Word or LibreOffice/OpenOffice. If your documents are in other formats, you can use these applications to convert them before the import. During import, QualiAnon copies the file(s) into the project folder. Your original files will not be modified.
- **Quit:** Close the application.

#### 3.1.2 Edit

- **Undo:** Undo the last operation.
- **Redo:** Redo the last operation that was undone. Undo and redo can be used to roll back and forth multiple times.
- **Find:** Search for text across all interviews in the project.
- **New Category:** Edit the category scheme used for anonymization in the project (see 2.2).
- **New Export:** Create a new export of the project documents with applied anonymization settings.

#### 3.1.3 Window

- **Show Message Log:** Display the log window at the bottom area.
- **Close Active Bottom Tab:** Close the currently showing tab on the bottom area.
- **Close All Bottom Tabs:** Close all the tabs displayed on the bottom area.

#### 3.1.4 Help

- **User Manual:** display the online version of this document in a browser
- **Qualiservice Website:** open [www.qualiservice.org](http://www.qualiservice.org)
- **QualiAnon @ Github:** open the related github page, <https://github.com/pangaea-data-publisher/qualianon>. Download the latest release, submit bug reports, discuss future enhancements to the tool.

### 3.2 Project structure tabs

On the left side there are three tabs containing the project structure.

- **Documents:** list with all documents that were imported into the project. A double click on an entry shows the document in the document area on the right.
- **Replacements:** list with all replacements in the project, sorted by category.
- **Exports:** list with all exports and detailed anonymization settings.

### 3.3 Document area

On the right side, the content of documents is displayed. This is the main area for working on anonymization. The area can hold multiple opened documents which can be switched by their tabs.

### 3.4 Bottom tab area

The bottom area can hold a variety of tabs, which help with anonymization. E.g. the message log, search results or a list of markers.

### 3.5 Status line

On the very bottom there is the status line. It displays the last application message on the left. On the right you can toggle the selection mode:

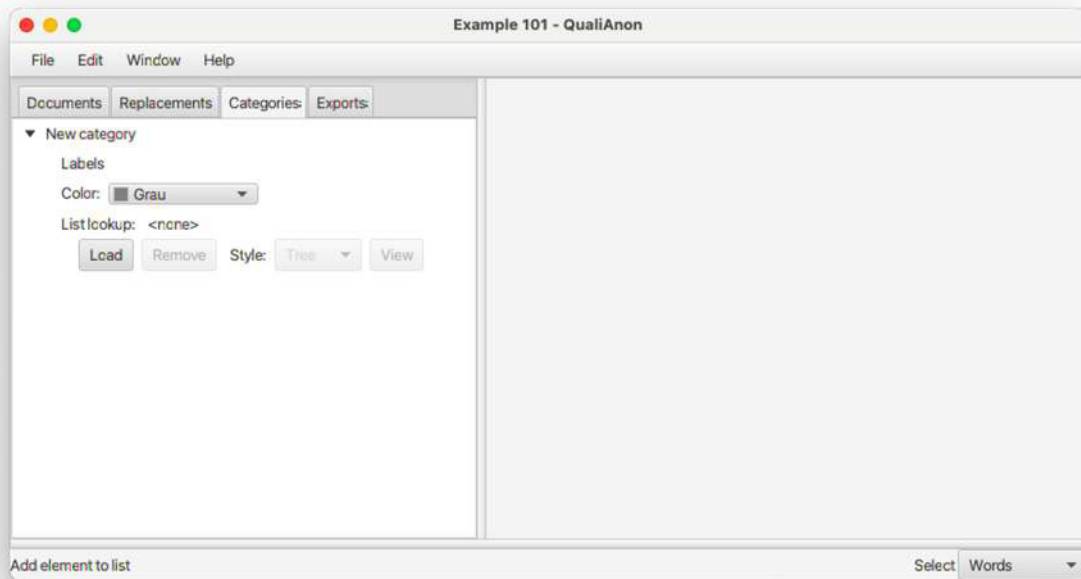
- **Words:** When selecting text in the document view, the selection is automatically extended to word boundaries. Single words can be selected with a single click.
- **Characters:** Selection is possible on the character level.

## 4 Create and edit categories

Categories are the classes of entities to be anonymized, the general schema of the anonymization. Depending on your project, you might want to set up different categories, e.g. Person, Location, Occupation, Disease.

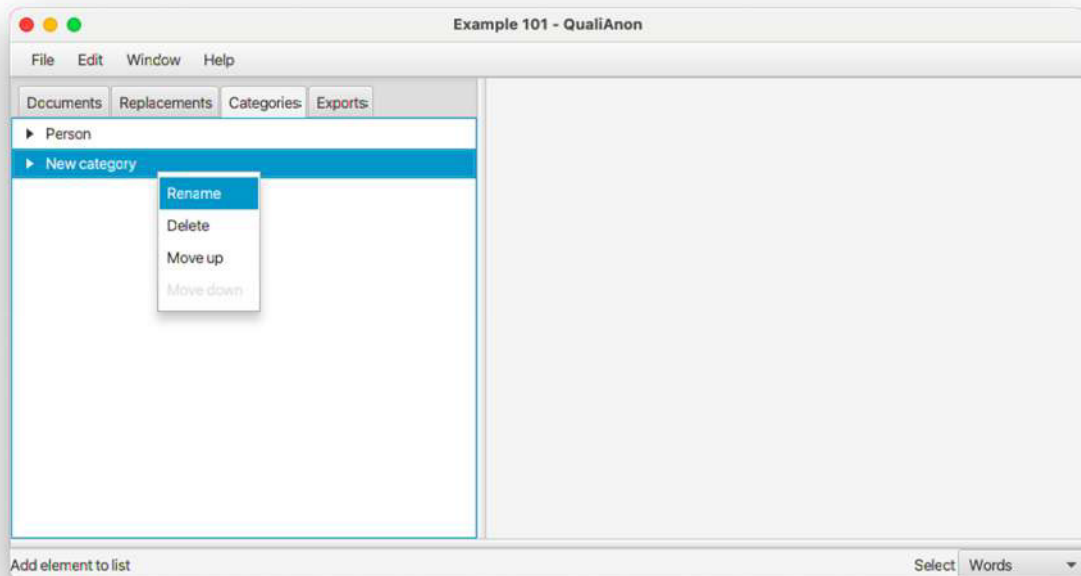
Categories can be assigned a color for display within documents. They comprise labels (free text or choice) to describe the anonymized entity on different levels.

## 4.1 Create a new category



1. Select from the menu: Edit -> New Category. This will create a new category and display the “Categories” tab.

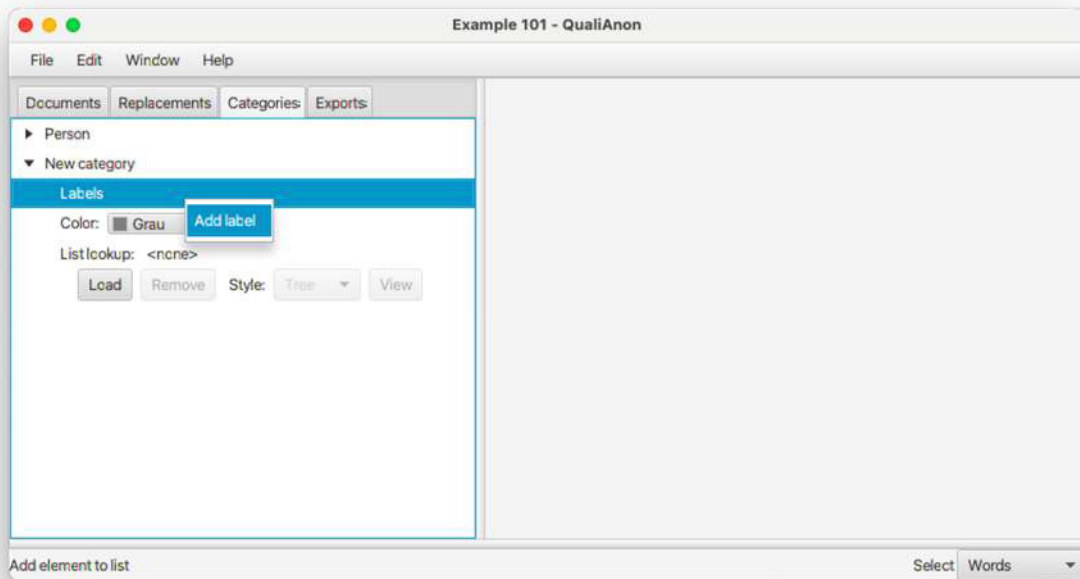
## 4.2 Organize categories



Right-click on a category name to open the context menu:

- **Rename:** change the name of the category, all usages will be updated
- **Delete:** remove the category and all usages
- **Move up / Move down:** change the order of categories

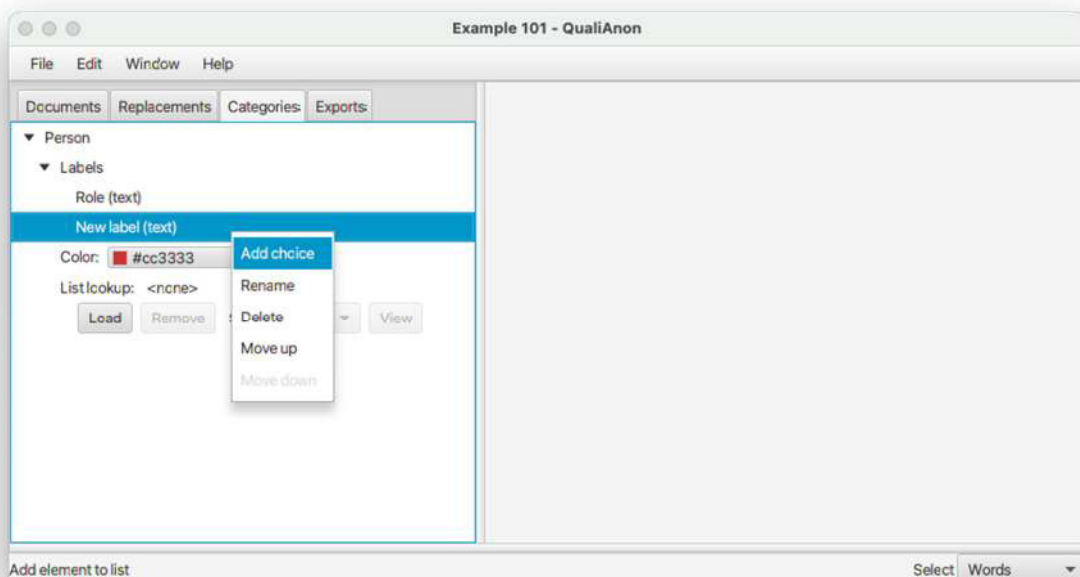
### 4.3 Create labels for a category



1. Right-click "Labels" within a category to open the context menu
2. Select "Add label"

There are two types of labels, either free text entry (text) or multiple choice from an enumeration (choice). New labels are always of type "text". To provide choices for a label, see next section.

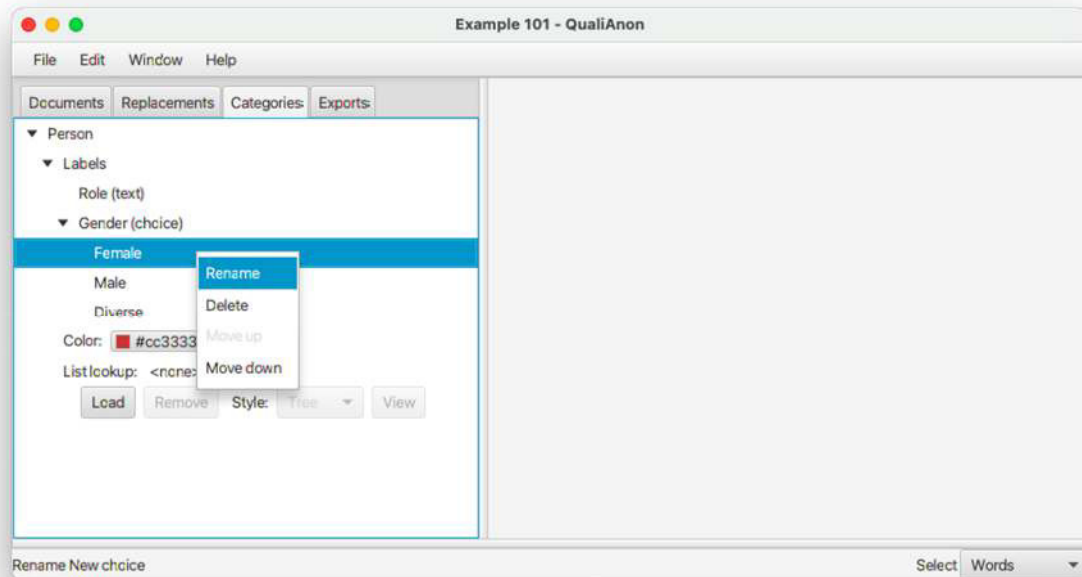
### 4.4 Edit labels of a category



Right-click on a label to open its context menu.

- **Add choice:** provide a value for the label. When adding the first choice to a label, it is converted from “text” to “choice” type. You can add multiple choices to each label to restrict the entry options.
- **Rename:** rename the label and its usages
- **Delete:** remove the label and its usages
- **Move up / Move down:** change the order of labels

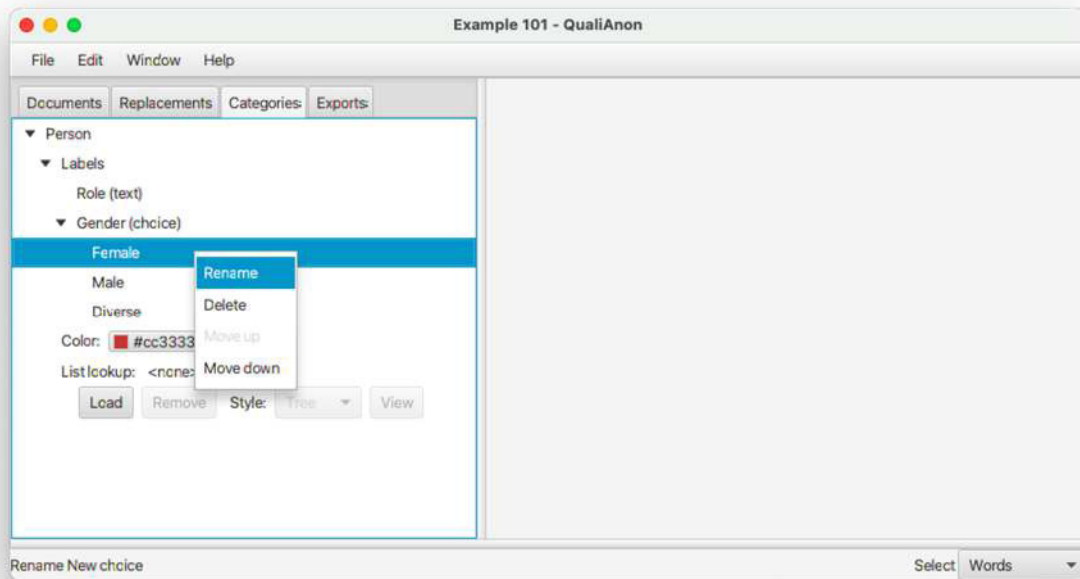
#### 4.5 Edit the choices of a label



Right-click on a choice to open its context menu.

- **Rename:** rename a choice (usages are not updated, yet)
- **Delete:** remove a choice from a label (usages are not updated, yet)
- **Move up / Move down:** change the order of choices

## 4.6 Load a list for label

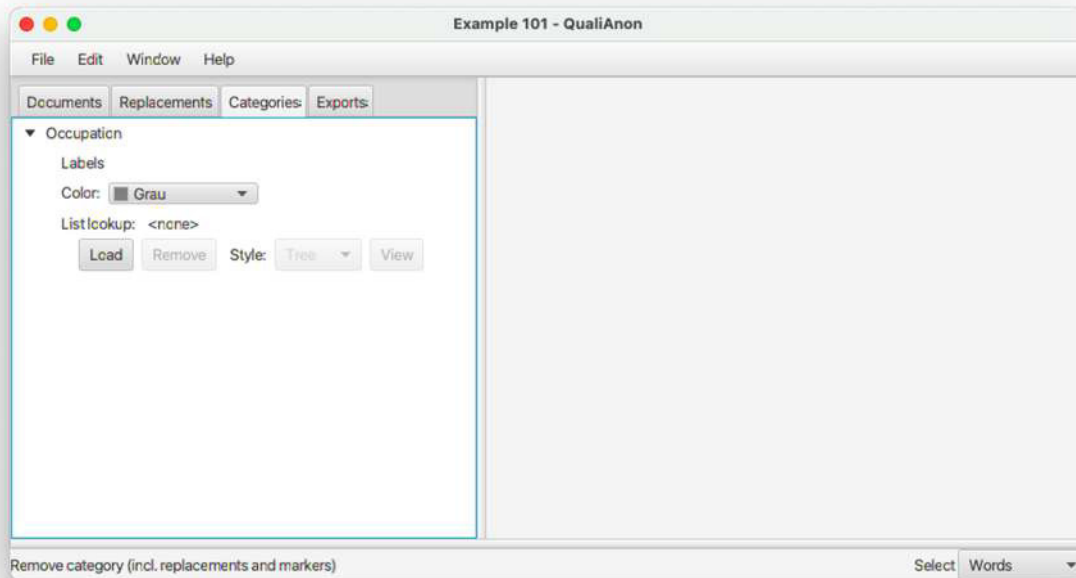


Right-click on a choice to open its context menu.

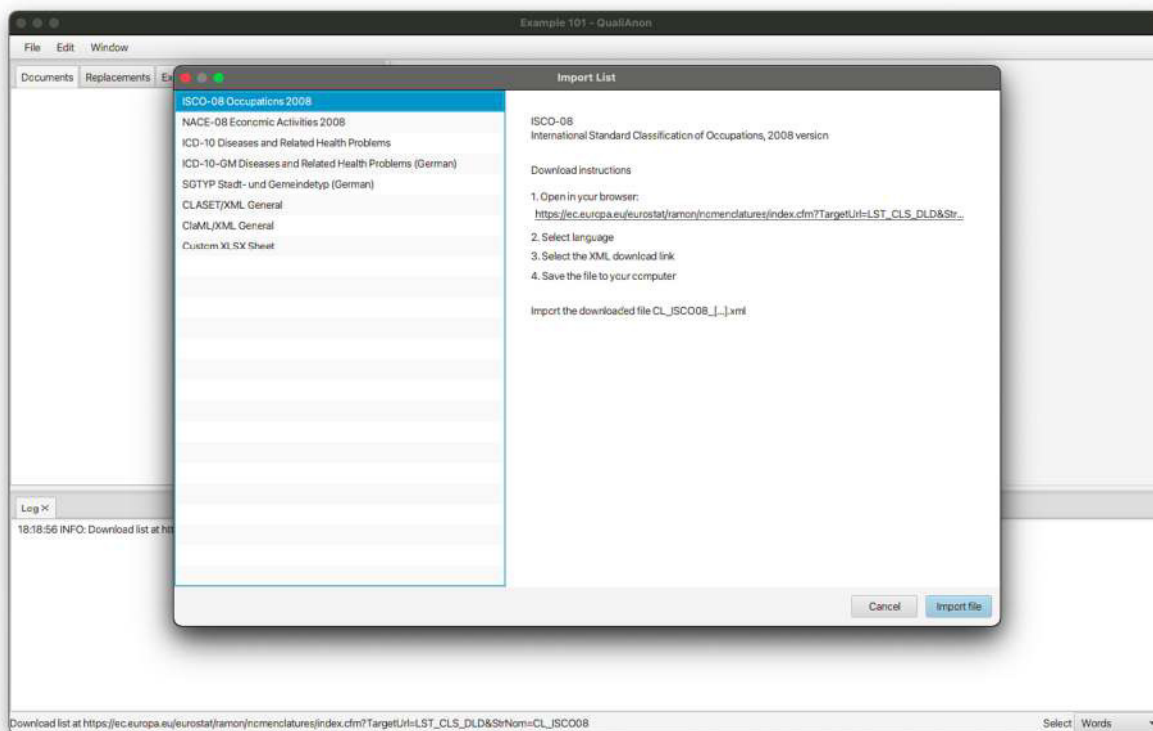
- **Rename:** rename a choice (usages are not updated, yet)
- **Delete:** remove a choice from a label (usages are not updated, yet)
- **Move up / Move down:** change the order of choices

## 4.7 Load a list for label lookup

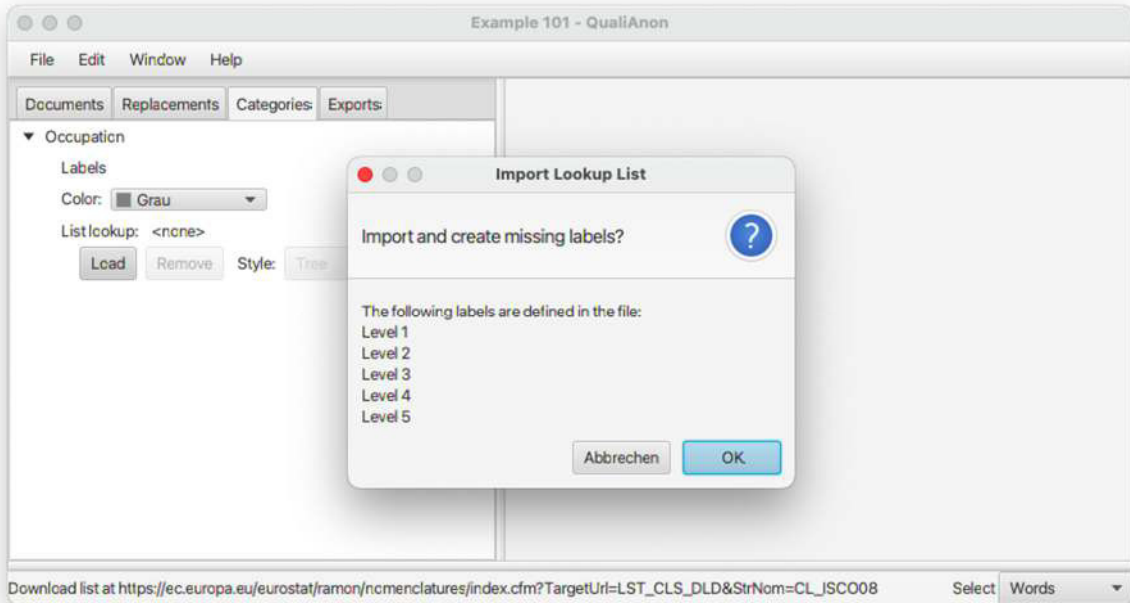
QualiAnon can load a number of standardized lists to fill labels automatically based on the anonymized words. It comes preconfigured with lists for occupations (ISCO-08) and economic activities (NACE-08) from the EU RAMON server, the classification of diseases (ICD-10) and a list of German cities and regions. It can also load custom lists from xlsx files.



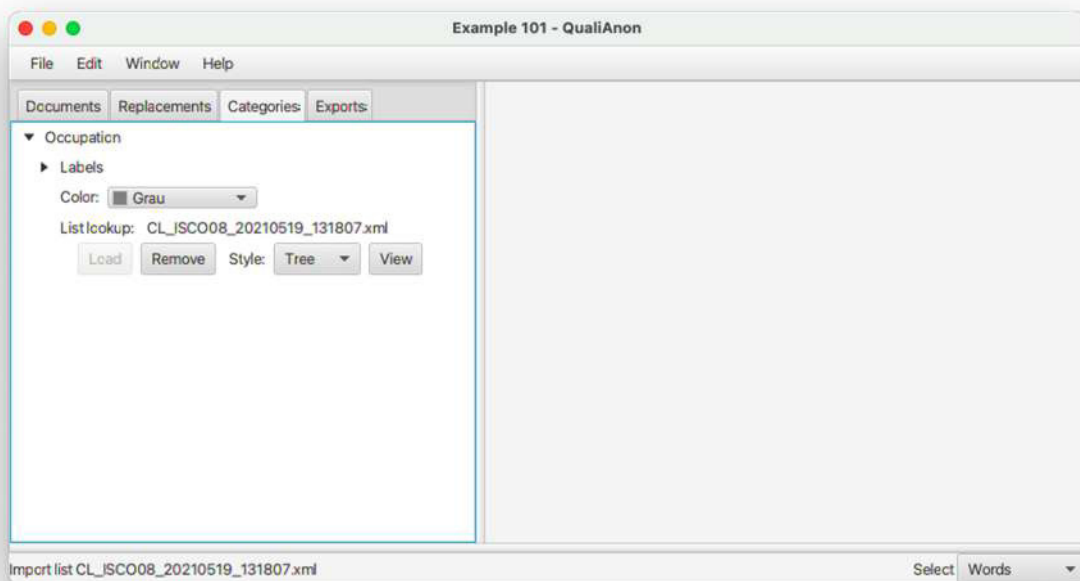
Press the “Load” button within a category to start the import.



1. Select a list type from the left.
2. On the right, you are given instructions on the download and preparation of list files.
3. Press “Import file” to select the prepared file and start the import.



4. Depending on the list, QualiAnon will create a number of labels.



5. Select a style: "List" is useful for non-hierarchical lists, like the list of German cities and regions. "Tree" is good for hierarchical lists, like ISCO, NACE, ICD.
6. Select "View" to preview the list content.
7. To remove a list from a category, select "Remove".

If your computer is disconnected from the Internet, you can use a different computer to download list files. Either copy the URL from the browser or message log into a text file and transfer it with a USB stick, or install QualiAnon on the connected computer. Then transfer the



downloaded files with a USB stick to the disconnected computer and continue with “Import file”.

It is possible to use lists in combination with additional labels, so that you can fill some labels from the list and add additional information manually. Just add more labels as described above.

#### 4.8 Using custom lists for replacement

You can easily use your own lists of replacements with QualiAnon.

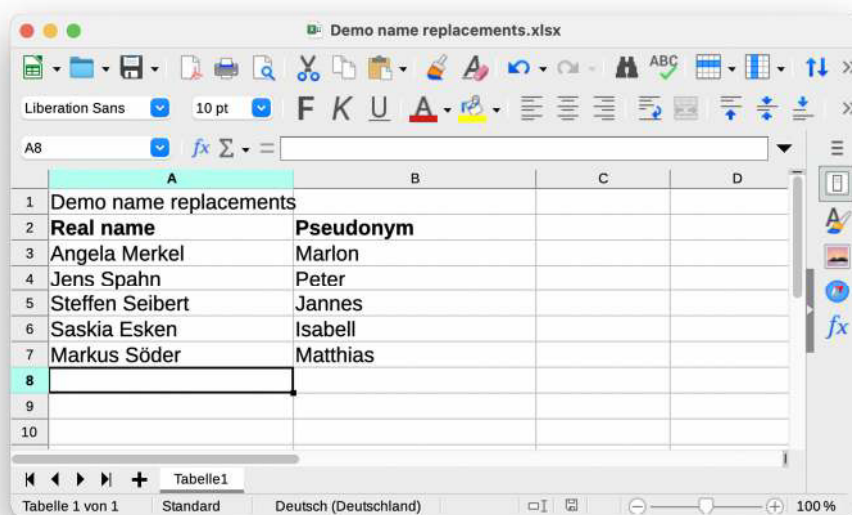
1. QualiAnon expects custom lists in the XLSX format with a simple format definition:

Cell A1: Citation information, source of data

Row 2: Header row

Rows 3 and following: Data rows

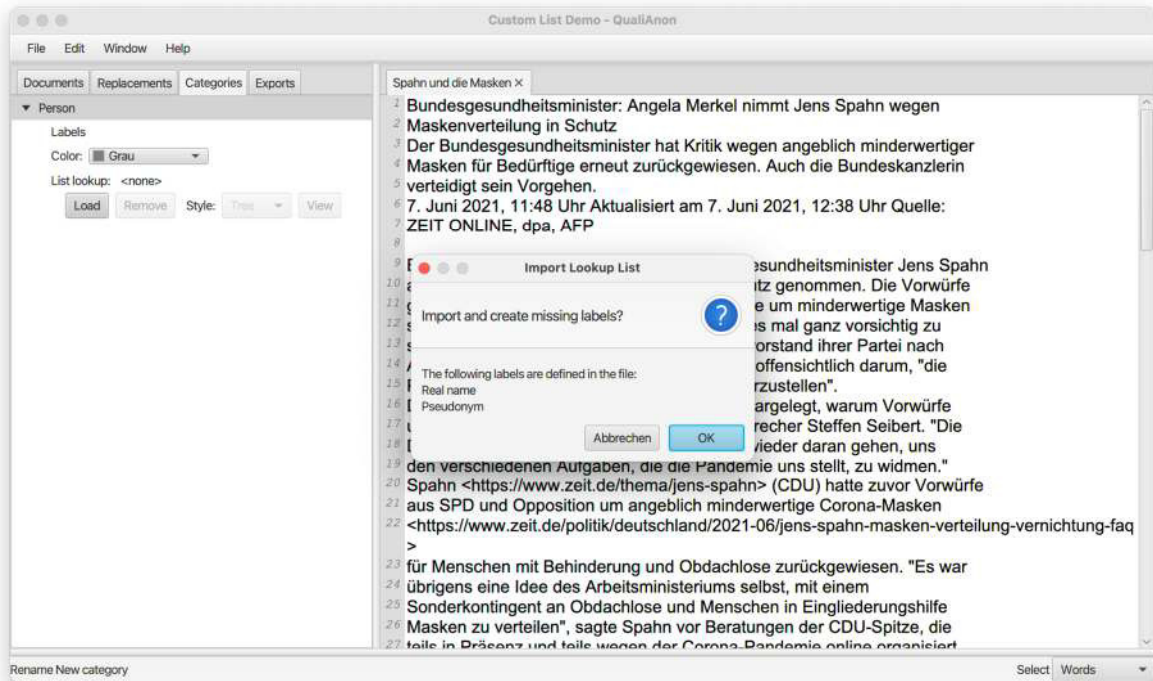
The following screenshot shows an example replacement table following this format.



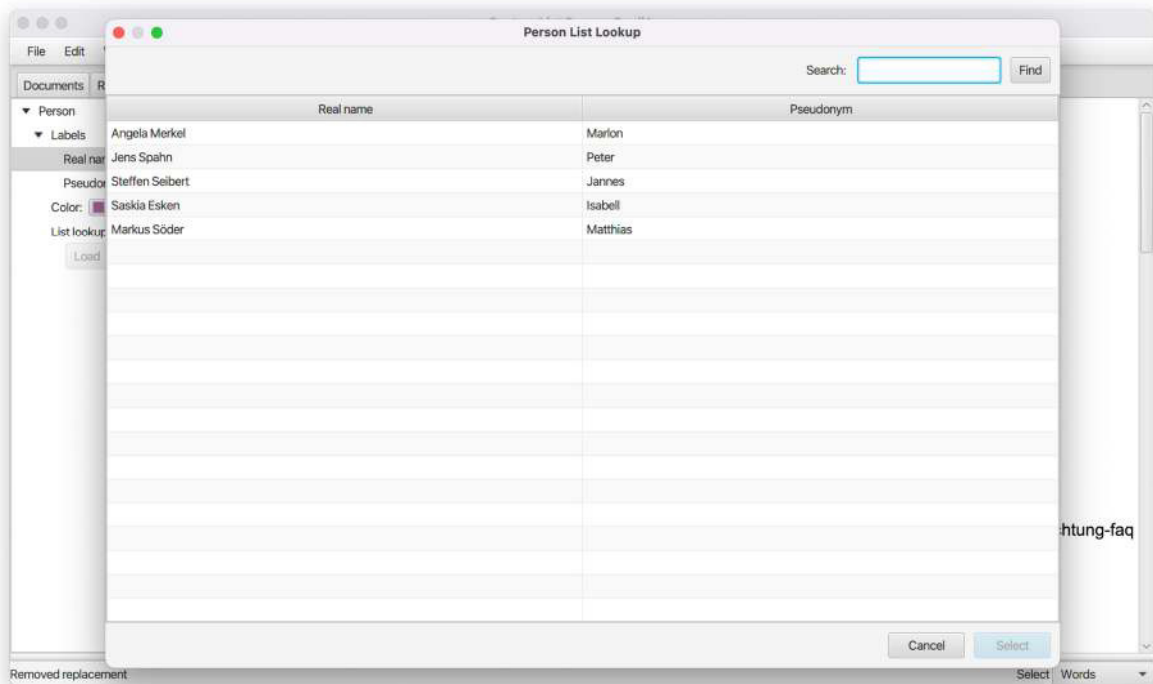
The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D
1	Demo name replacements			
2	<b>Real name</b>	<b>Pseudonym</b>		
3	Angela Merkel	Marlon		
4	Jens Spahn	Peter		
5	Steffen Seibert	Jannes		
6	Saskia Esken	Isabell		
7	Markus Söder	Matthias		
8				
9				
10				

2. Save the XLSX file outside of the QualiAnon project folder. During import, QualiAnon will copy the file into the project.
3. In QualiAnon, create a new category or use an existing category for your replacement list (see section 4).
4. Import your custom XLSX file as described in section 4.6. Select “Custom XLSX Sheet” when asked for the kind of list.
5. QualiAnon will read the file and offer to create a label for each column in the file with the name given in the header row of the XLSX file.



6. Verify the result of the import with the “View” button on the category.



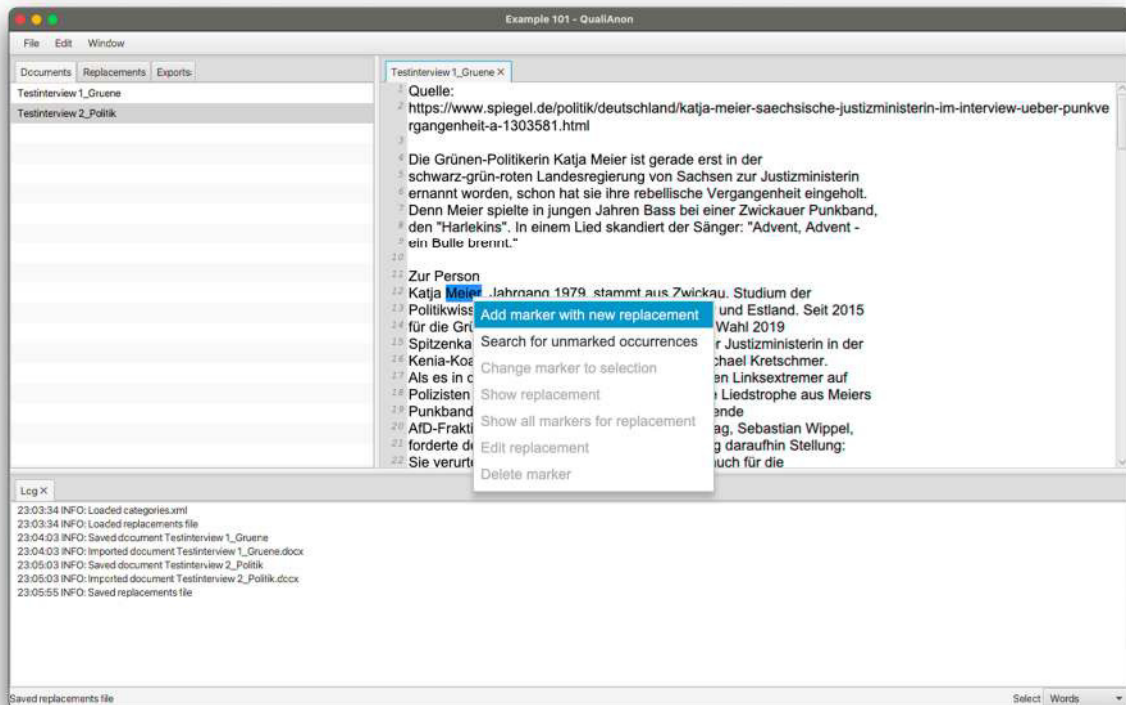
You can now start to create replacements supported by your list, as described in section 5.

## 5 Working with replacements and markers

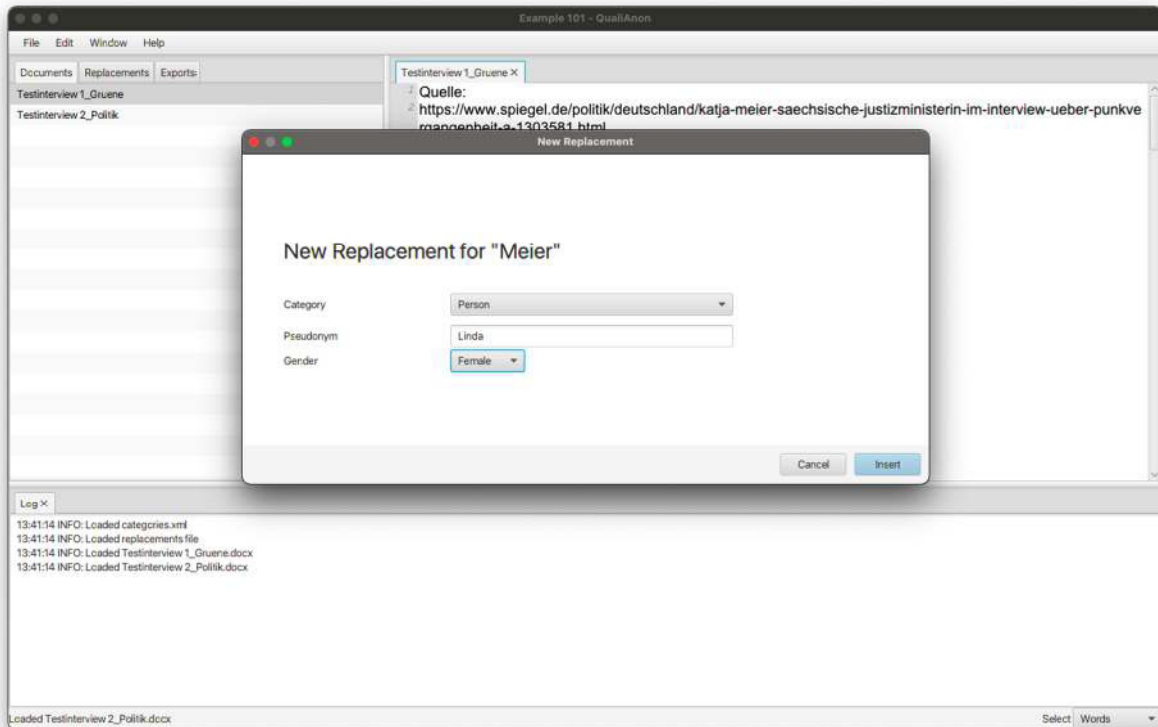
In QualiAnon, replacements are distinct entities to be anonymized, e.g. a specific person or a specific location. A replacement is of one category and can be annotated with the labels defined within the respective category.

A marker is an occurrence of a replacement within a document. E.g. a person (represented by a replacement) occurs in a document several times with different words. Sometimes the full name may be mentioned, sometimes only the last name. Each occurrence has a separate marker which is connected to the same replacement.

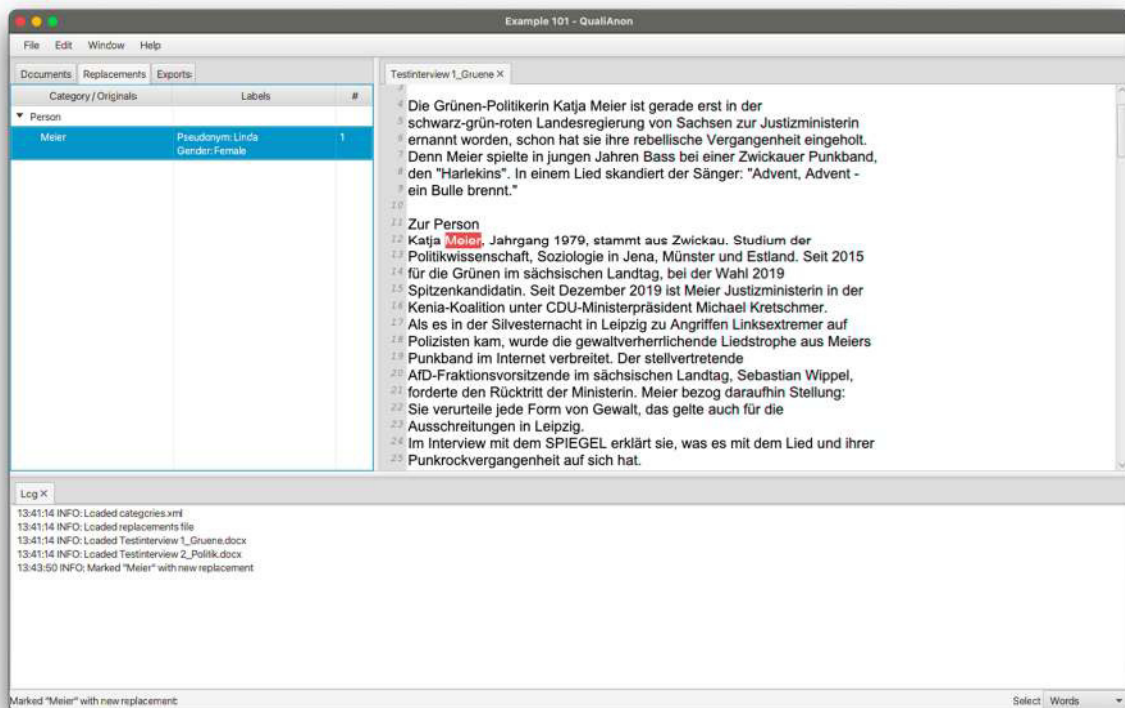
## 5.1 Add a marker with a new replacement



1. Select a word or phrase in a document.
2. Open the context menu with a right mouse button click and select "Add marker with new replacement"



3. Select a category and fill in the labels.
4. Press "Insert"

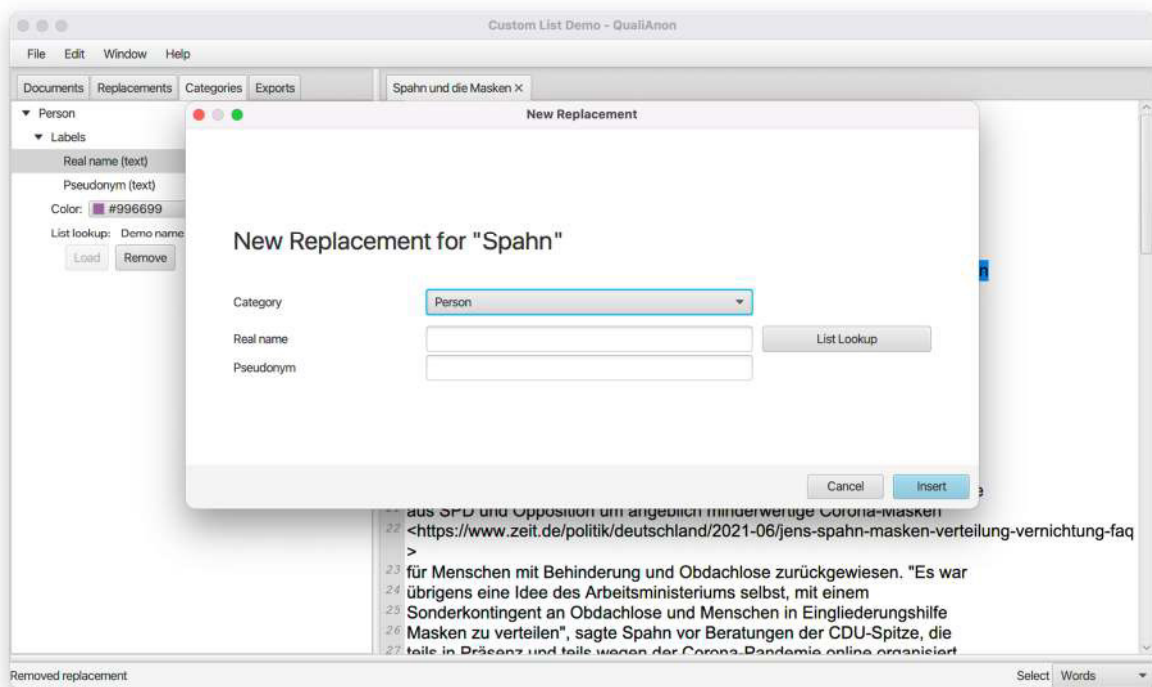


The marker is now inserted into the document, highlighted in the color of the category. The related replacement can be inspected in the replacements tab.

## 5.2 Populate labels from a replacement list

If a category is supported by a replacement list, you do not have to fill the labels by hand.

1. For categories with a replacement list, you have the option to do a "List Lookup".

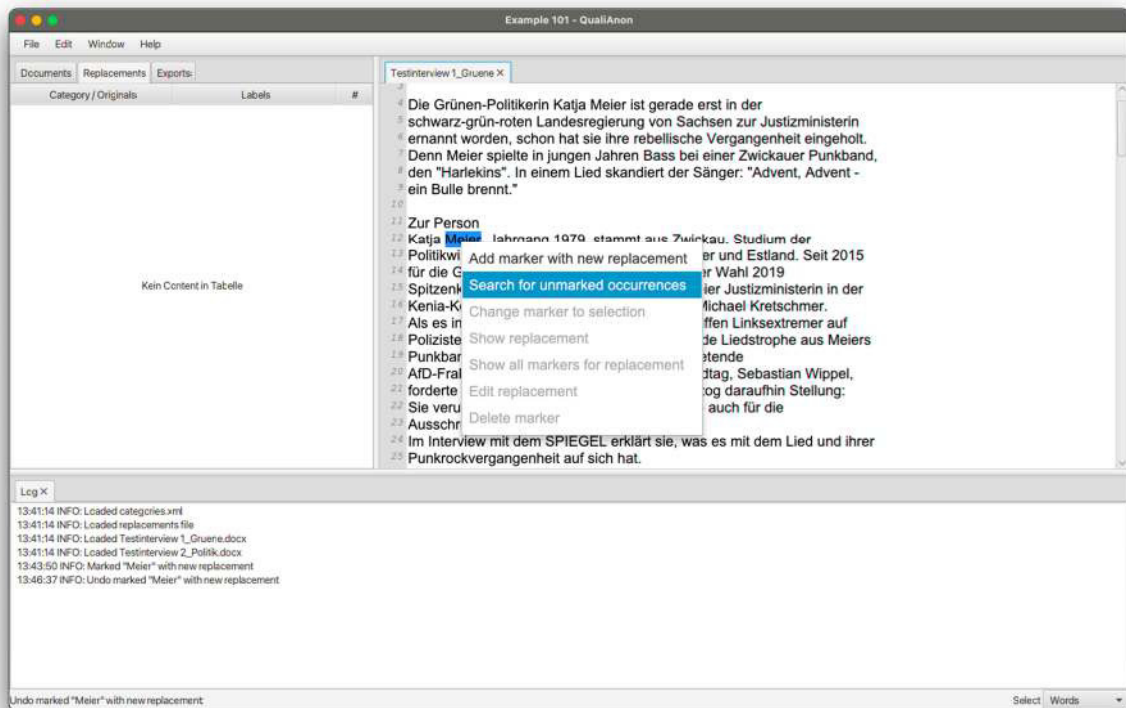


2. The list is filtered by your current selection in the text. You can change the filter in the top right search field.

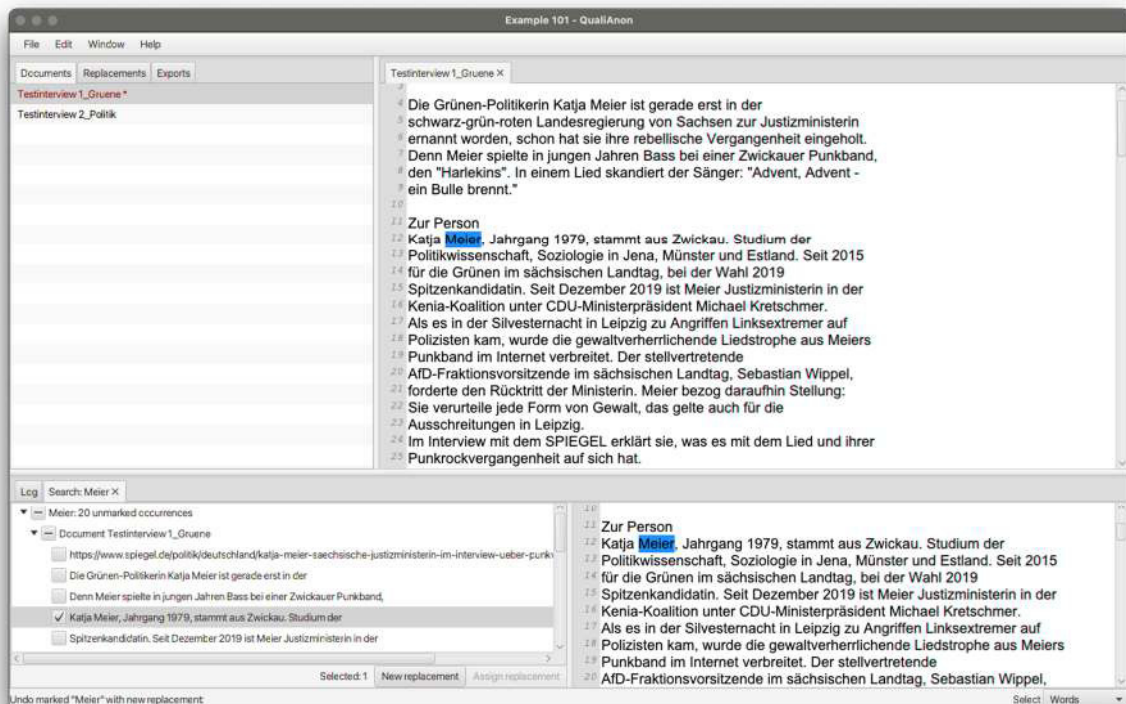




### 5.3 Mark multiple occurrences of a word or phrase with a new replacement



1. Select a word or phrase in a document.
2. Open the context menu with a right mouse button click and select “Search for unmarked occurrences”

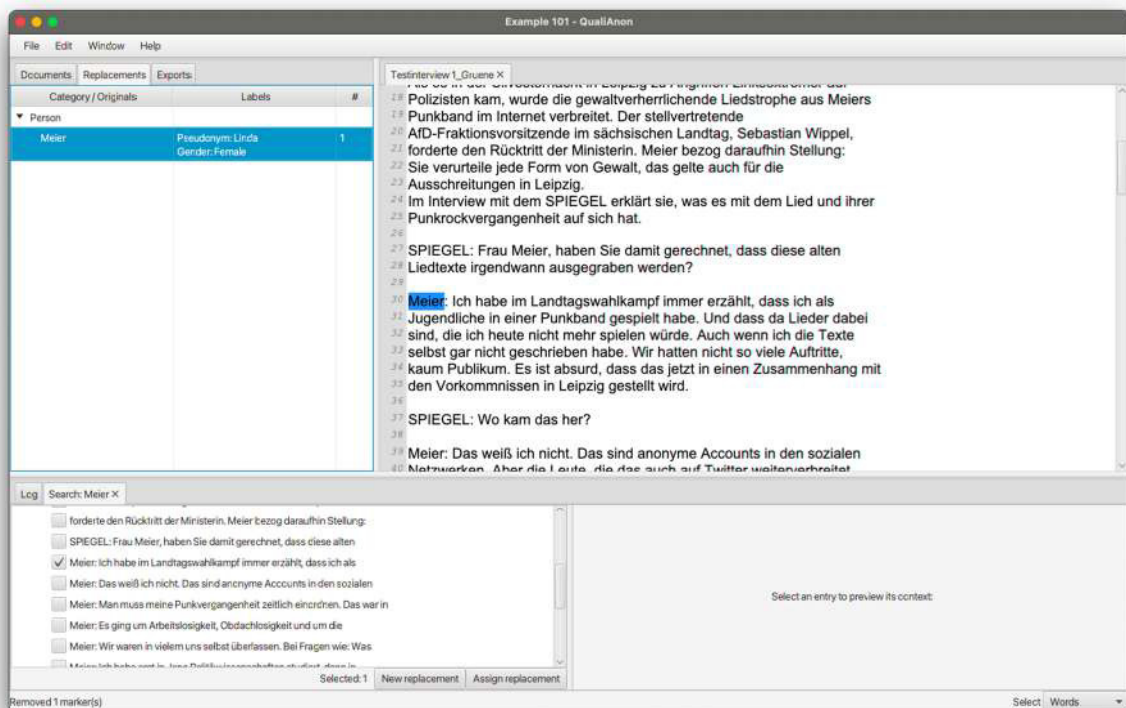


- The search tab at the bottom shows all occurrences of the searched word or phrase over all documents in the project on the left side. On the right, you can inspect the context of each occurrence. Review all occurrences and set the checkboxes on the left side to include them for marking.
- Press “New replacement” at the bottom of the search results to mark all checked occurrences and create one new replacement (see section 4.1, step 3)

#### 5.4 Add markers to an existing replacement

If you have already created a replacement for an entity, all markers have to be assigned to this replacement. You should not have multiple replacements for the same entity. Otherwise you will run into trouble when exporting with the “counting” option and the data in the project will probably be inconsistent.

- If you find another occurrence for an already created replacement, follow steps 1-3 of section 4.2.

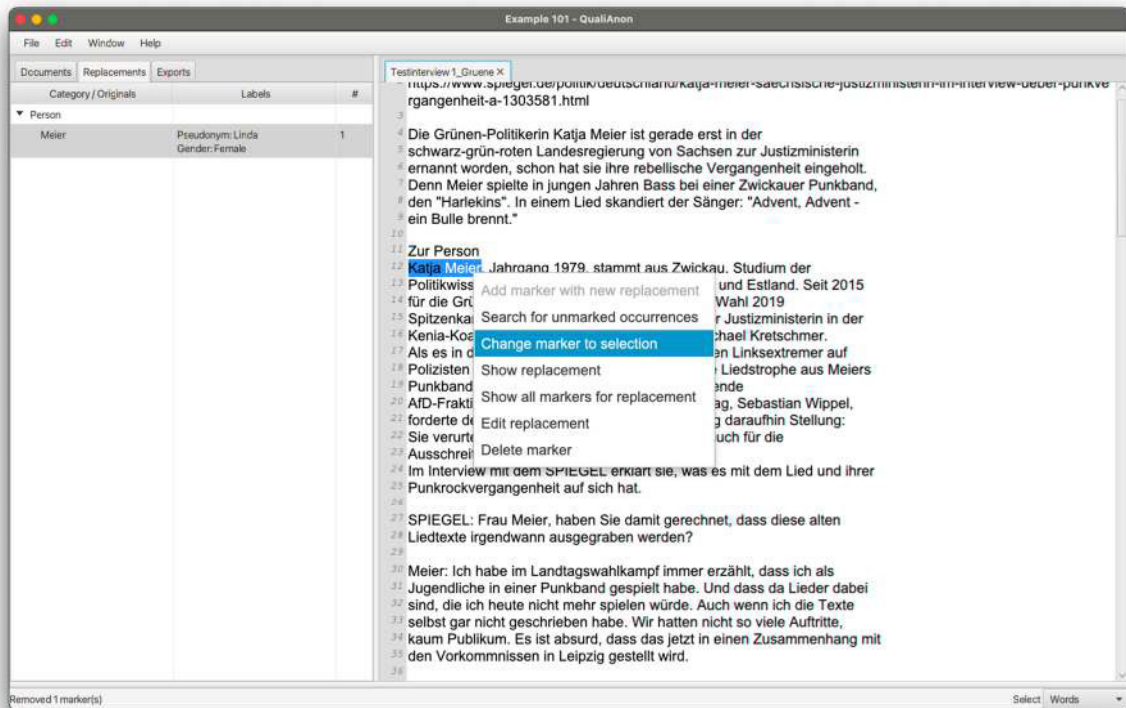


- Click on a replacement in the replacements tab to select it as a target for the new occurrences.
- Press “Assign replacement” at the bottom of the search results to set markers related to the selected replacement.



## 5.5 Change marker to selection

In some cases, the markers you set might not comprise the whole entity. E.g. you mark the last name “Meier”, but there is a case where the full name is given in the document.

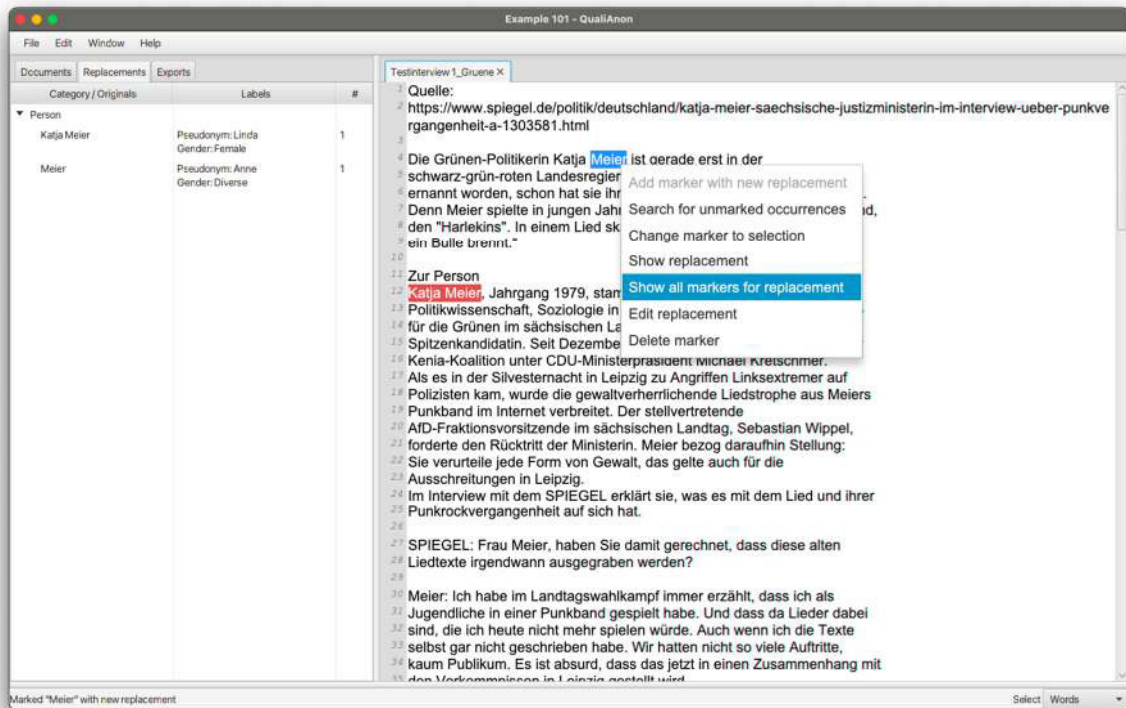


To change the extend of a marker:

1. Select the correct range of words or characters. A part of the existing marker has to be within the selection for this to work.
2. Select “Change marker to selection” from the context menu.

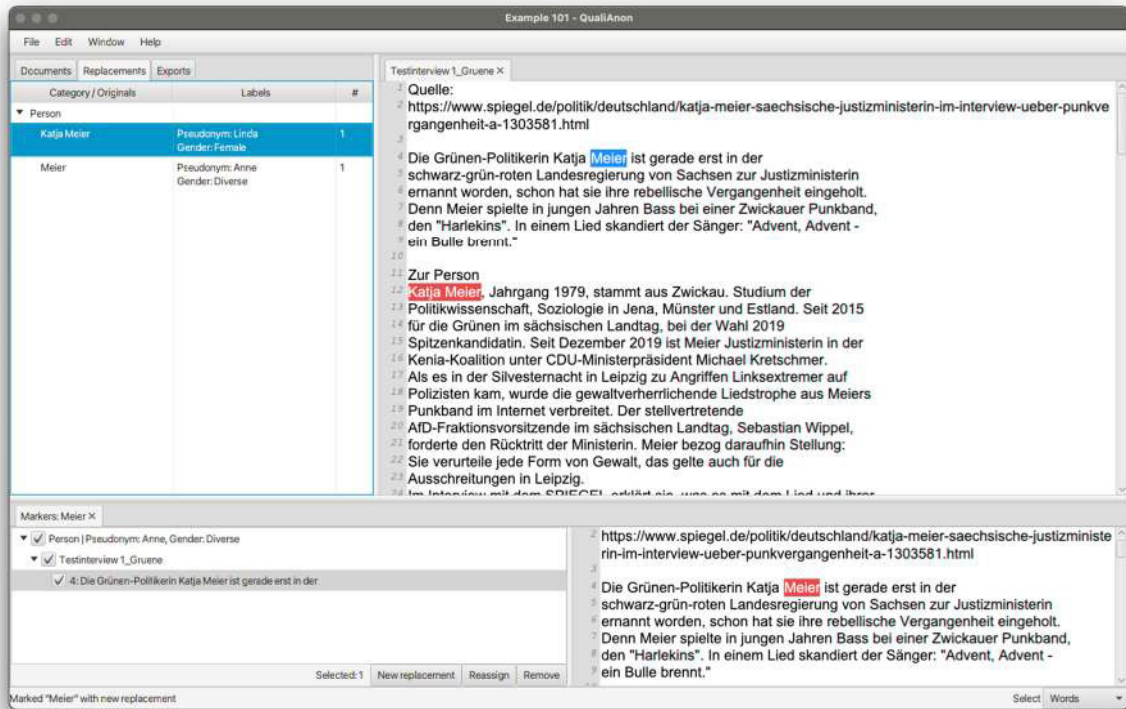
## 5.6 Reassign markers to another replacement

If you have assigned markers to wrong replacements, or have created multiple replacements for the same entity, you can easily fix it.



In this example, there are separate replacements for the same entity in the document (person Meier).

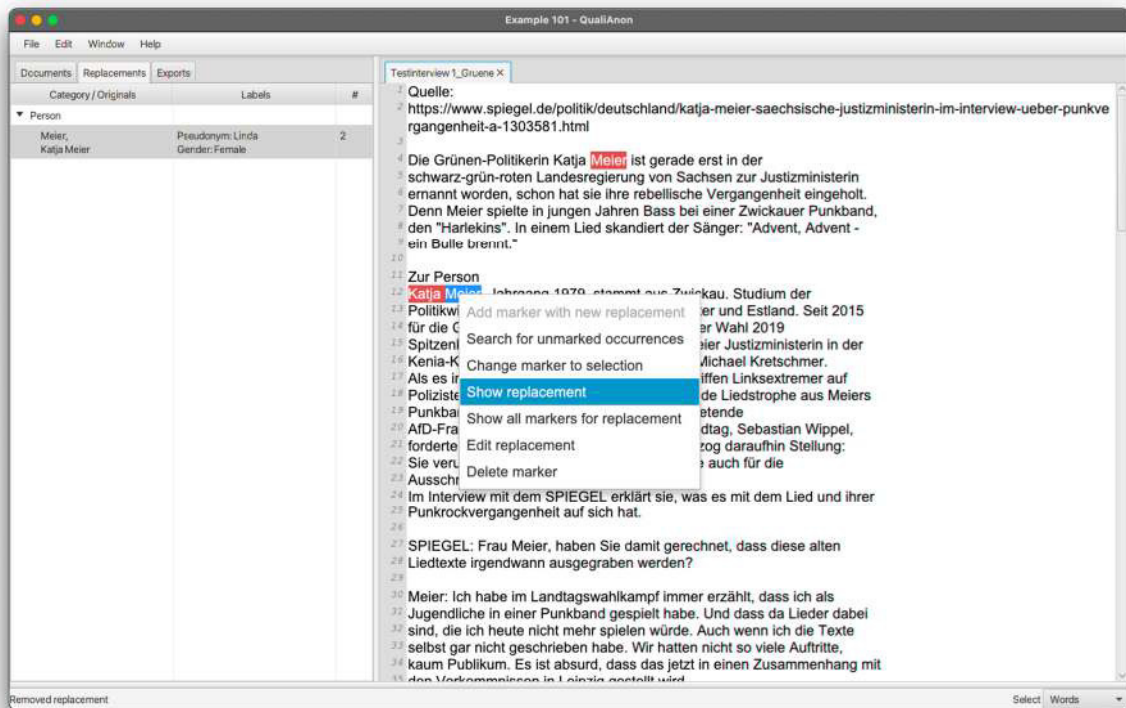
1. Select a marker to be reassigned.
2. Select "Show all markers for replacement" from the context menu.



3. A markers tab opens at the bottom. Review and check all markers that you want to reassign to another replacement.
4. Click on a replacement in the replacements tab to select it as the target.
5. Press "Reassign" at the bottom of the markers tab.

Alternatively, you can right click on a replacement in the replacements tab and select "Show markers" from the context menu. Then resume with step 3.

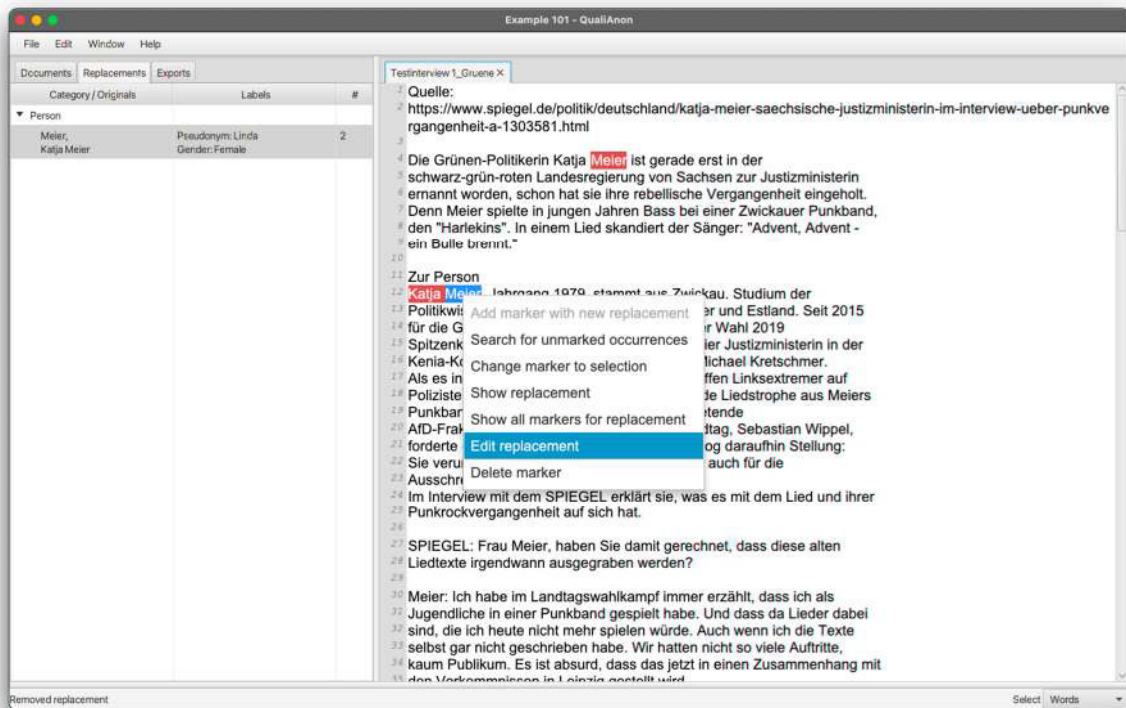
## 5.7 Find replacement and label information for a marker



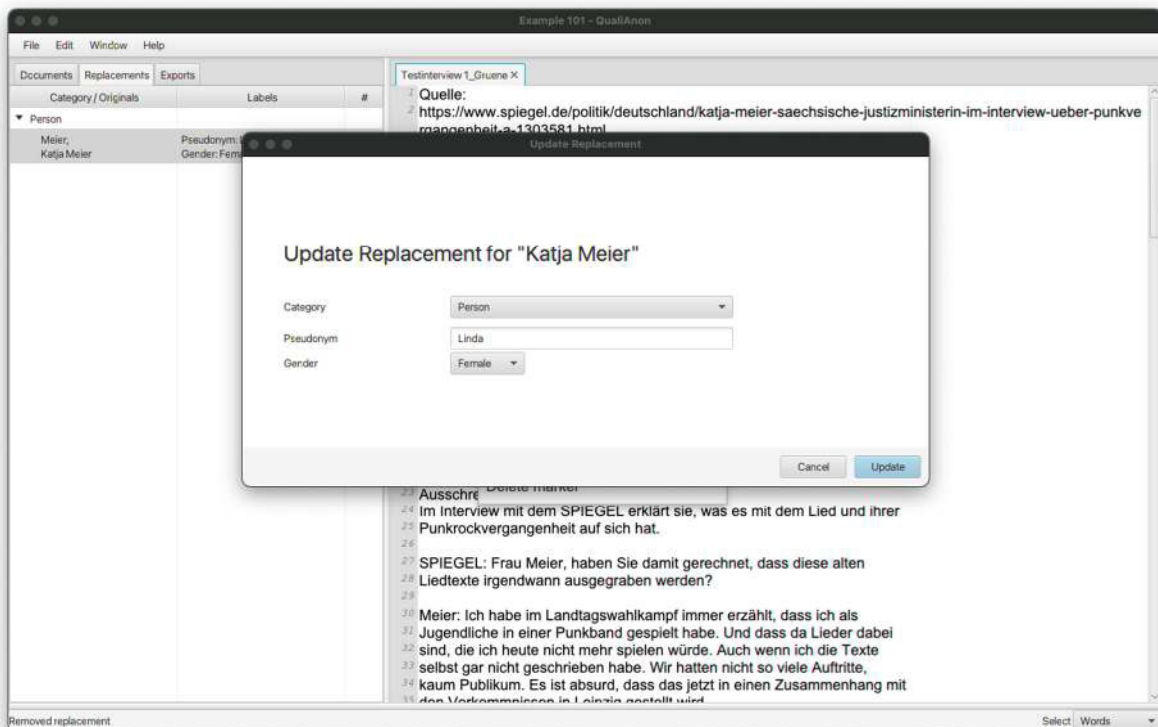
1. Select a marker in a document
2. Select "Show replacement" from the context menu

QualiAnon shows the replacements tab with the corresponding replacement selected.

## 5.8 Edit a replacement



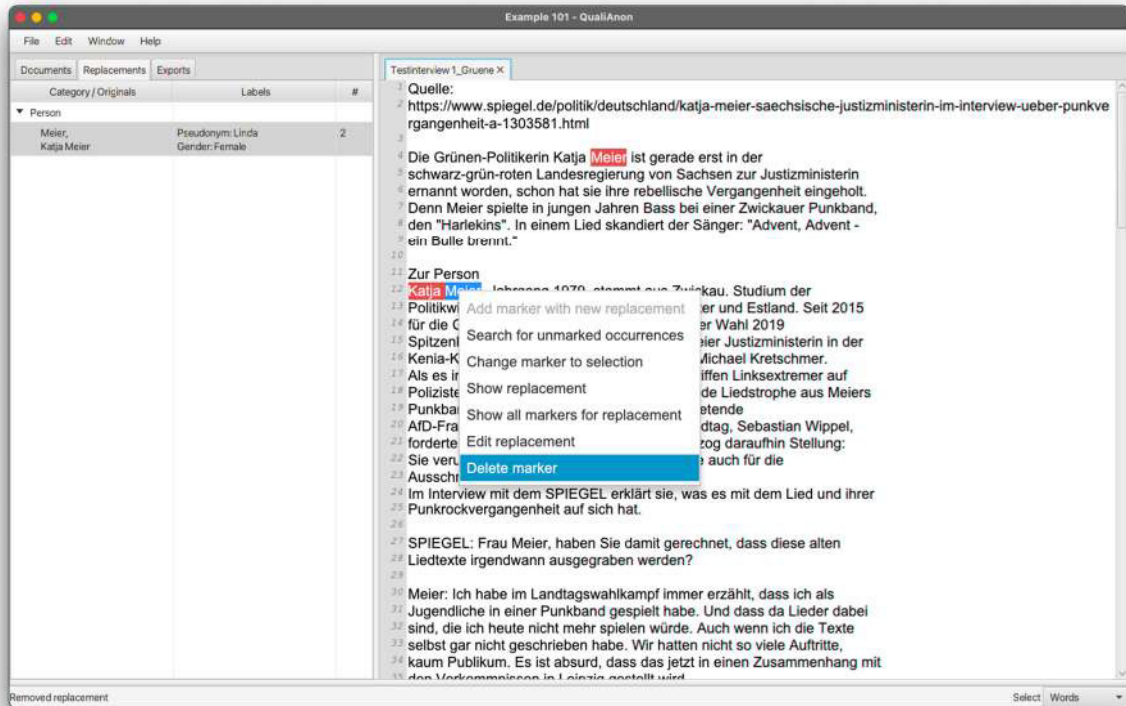
1. Select a marker
2. Select "Edit replacement" from the context menu



3. Edit the labels and category of the replacement in the dialog

Alternatively, you can right click on a replacement in the replacements tab and select “Edit” from the context menu. Then resume with step 3.

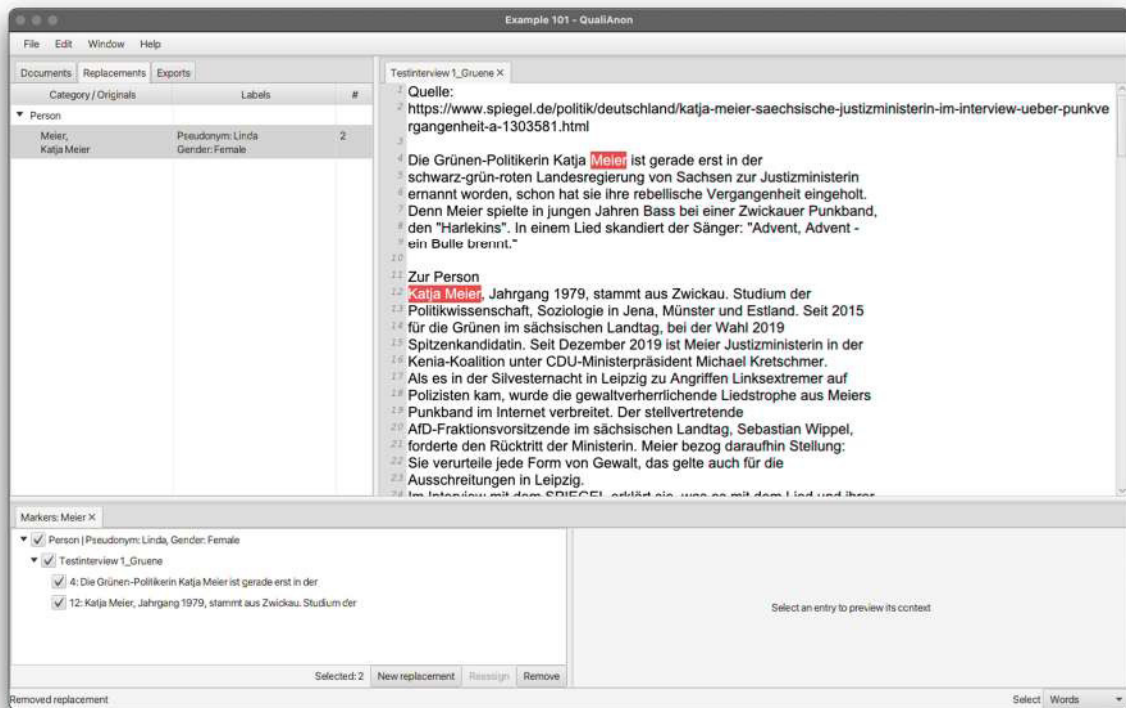
## 5.9 Remove a marker



1. Select a marker
2. Select “Delete marker” from the context menu

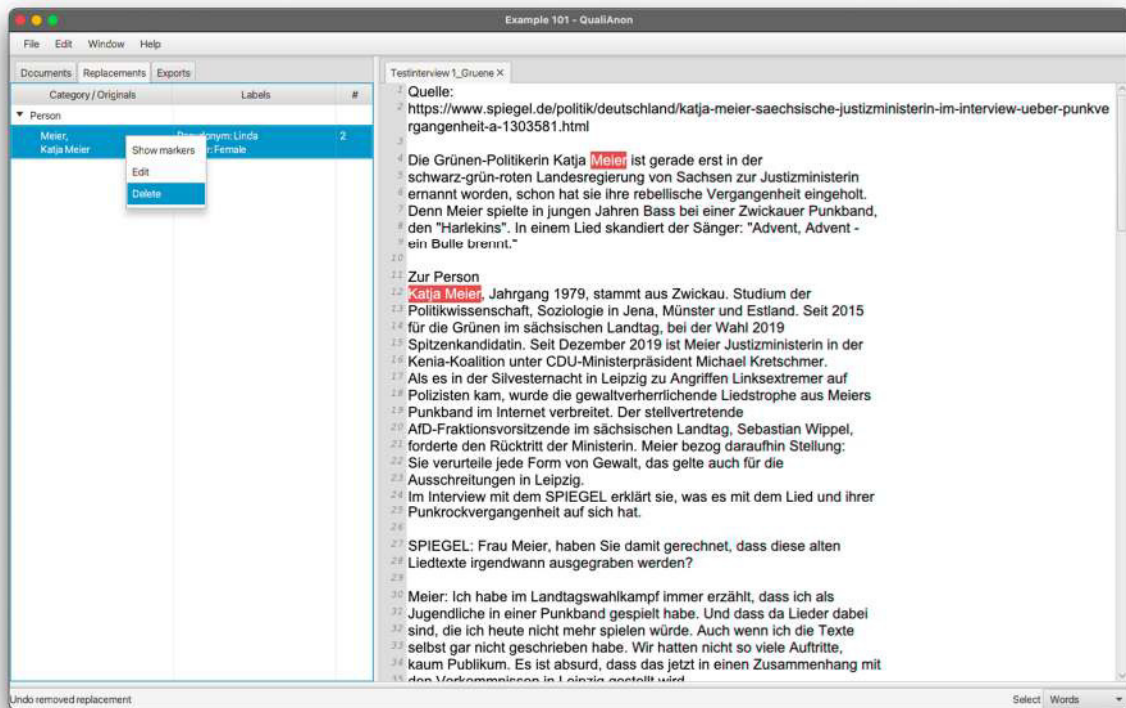


## 5.10 Remove multiple markers



1. Use the "Show all markers for replacement" function, as explained in section 4.5.
2. In the bottom marker tab, review and check all the markers you want to remove
3. Press "Remove" at the bottom of the marker tab

## 5.11 Remove replacement with all related markers

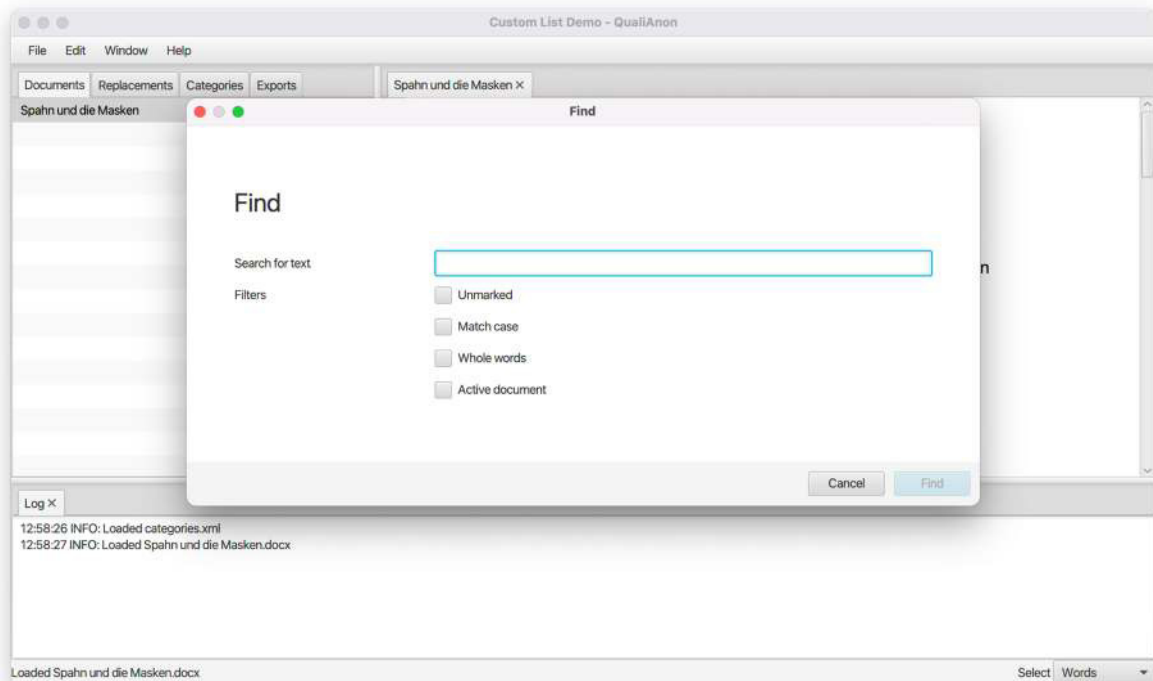


1. Select a replacement from the replacements tab
2. Select "Delete" from the context menu



## 6 Find text

QualiAnon can search through all your documents within a project. Select from the menu: Edit -> Find

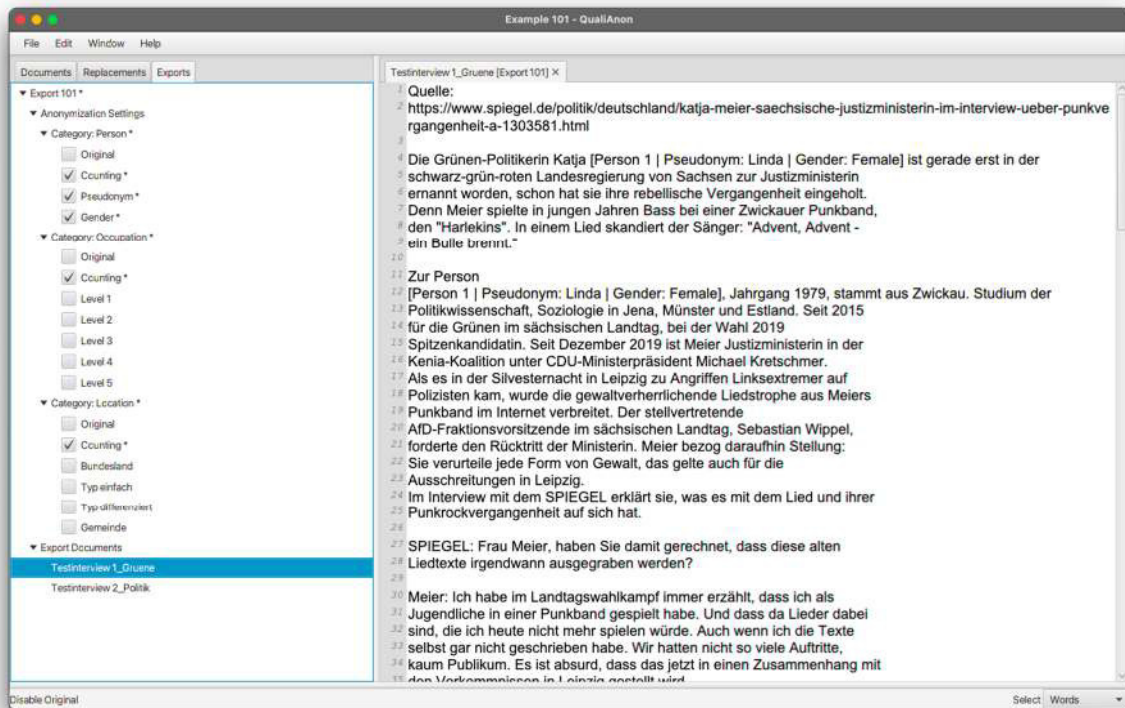


- **Search for text:** the text you are looking for
- **Unmarked:** limit results to text that has no markers
- **Match case:** limit results to text with exactly the same capitalisation as in your search text
- **Whole words:** limit results to occurrences of whole words and omit occurrences that are only part of a word
- **Active document:** only search the active document and not all documents in the project

## 7 Creating and configuring exports

When QualiAnon exports the documents of a project, it replaces each marked word or phrase with the information given in the related replacement. Since the labels in their entirety may contain too much information, you can configure which labels to include in an export.

1. Create a new export from the menu: File -> New Export ...
2. Choose a name for the export

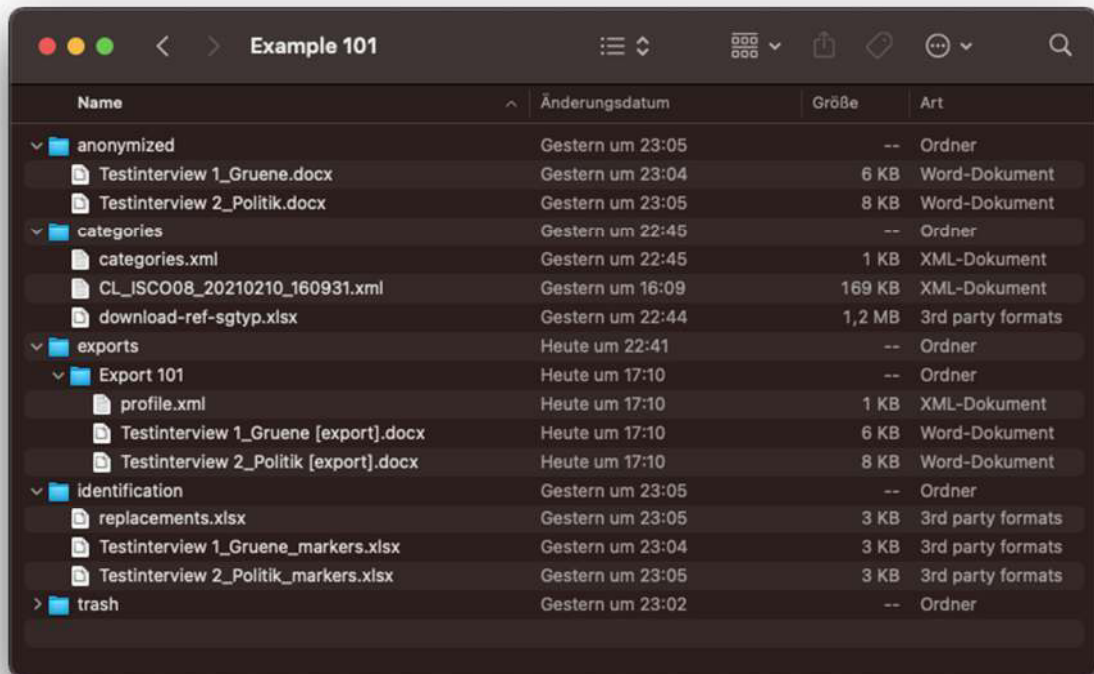


3. For each category, you can configure, whether
  - a. to display the original without any anonymization,
  - b. to use counting to automatically enumerate the anonymized entities, and
  - c. which labels to display.
4. Double click the entries under the “Export Documents” section to preview the export of the documents of the project.
5. When you are finished, open the context menu on the export item and select “Save settings & documents”

## 8 Project directory and files

The project folder is organized into different directories and files.

### 8.1 Directory layout



- **anonymized:** this folder contains the documents of the project with all marked text removed. Each marked word or phrase is replaced by an identified, e.g. [M:76ec5e26-a314-40ad-9a0e-566e4273b613:].
- **identification:** this folder contains the missing information from the documents in “anonymized”. For each document, an xlsx table is saved which lists all marker identifier together with the original value and another identifier, which references the replacement. The file replacements.xlsx contains all these replacements with all the label values entered during anonymization.
- **categories:** this folder contains the category schema (categories.xml) and the imported list files.
- **exports:** for each export there is a subfolder containing the export settings (profile.xml) and the exported documents.
- **trash:** this folder stores deleted documents and exports.

### 8.2 Get the exported documents

After exporting (see section 5), the exported documents are stored in the export folder. Open the docx files to verify that they are correctly anonymized.

### 8.3 Reuse categories for another project

You might want to use the same category configuration for multiple projects. Copy the categories folder with its content from one project to another. Reopen the project in QualiAnon, so that the configuration is loaded.

### 8.4 Archive a project securely

To increase security, you can store your project separated at different locations. You need two independent storage locations.

1. Move the identification directory to location a)
2. Delete the trash folder
3. Delete the exported documents, but not the profile.xml files in the same directories
4. Move the folders anonymized, categories, and exports to location b)

### 8.5 Remove the originals, but keep the labels

You can open the xlsx files in the identification folder to remove sensitive information altogether. E.g. you might want to remove the originals from the [...]\_markers.xlsx files. First, make a backup of the files you want to edit. Open the file, remove the data and save. Then, reopen the project in QualiAnon. QualiAnon can still work with labels, but can no longer display the originals.

## 9 Bug reports and feature requests

Join the QualiAnon github project to report bugs and request new features. Currently, the repository is not openly accessible. Contact us at <https://www.qualiservice.org/en/contact.html> to gain access.

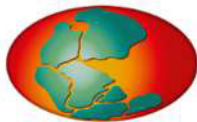
<https://github.com/pangaea-data-publisher/qualianon/issues>

## About Qualiservice.

The research data center Qualiservice archives qualitative social science research data from various disciplines and makes them available for scholarly reuse. Our services are secure, flexible, and research-oriented. They include personal and study-specific consulting, curation and preparation of your data for reuse and long-term archiving as well as the provision of archived research data and relevant contextual information. Internationally connectable metadata make datasets searchable and findable. Persistent identifiers (DOI) ensure sustainable citability of data and study contexts.

In June 2019, Qualiservice was accredited by the RatSWD and follows its criteria for quality assurance. Qualiservice is committed to the DFG guidelines for ensuring good scientific practice and also takes into account the FAIR Guiding Principles for Scientific Data Management and Stewardship and the OECD Principles and Guidelines for Access to Research Data from Public Funding.

## Das Qualiservice-Konsortium:



PANGAEA – zertifiziertes  
Weltdatenzentrum



SOCIUM – Forschungszentrum  
Ungleichheit und Sozialpolitik



Staats- und  
Universitätsbibliothek  
Bremen

Staats- und  
Universitätsbibliothek  
Bremen



GESIS – Leibniz Institut für  
Sozialwissenschaften



Fachinformationsdienst Sozial- und  
Kulturanthropologie (FID SKA)



Gefördert von

