



MQIC

MAXQDA INTERNATIONAL CONFERENCE

Spotlight: Management of large data sets in MAXQDA

Christian Schmieder, Ph.D.

Assistant Director for Data Governance, University of Wisconsin-Madison, Division of Extension

<https://fyi.extension.wisc.edu/datajams/>

23. Februar 2024

Intro



TUBS - This SVG [locator map](#) includes elements from this [locator map](#):
Location of state of Wisconsin in United States in the [United States CC BY-SA 3.0](#)
https://en.wikipedia.org/wiki/Wisconsin#/media/File:Wisconsin_in_United_States.svg



Extension
University of Wisconsin-Madison



Intro

- 600 educators & faculty
- 70+ counties and tribal nations
- 3500+ narrative impact reports/yr



Key challenges with large datasets

- Manual workflows become unfeasible
- Manual activation of data become unfeasible
- Dataset size needs to be managed
- Requires more attention in terms of process documentation
- External files (PDFs) become unfeasible

Overview

- Key task: Cleaning & Import
- Key task: Structuring Data in MAXQDA
- Key task: Bulk Movement & Deletion of Data
- Key task: Documentation

Cleaning Data: An iterative task

- Clean data outside of MAXQDA
- Make plenty of test imports
- Run mock analysis and data structuring processes ahead of time
- Keep notes!

Import from Excel whenever you can

Document Group	Document Name	Abstract Sentence	Type of activity	Outcome Statement/Supplemental Information	Report Year	Hor
First Nations +	4829	A series of presentations for Native Americans where they learn how cultural knowledge and water science can connect on water topics. This effort was designed to increase Native American confidence to be local advocates for water.	Programming/work that occurred/is occurring	Tribal Nations have been stewards of the local water resources in the area for a very long time. Cultural knowledge gained from this long relationship with the nature, also known as traditional ecological knowledge, is a needed perspective when addressing water issues. Roberts created and delivered a series of	2021	Was
		A one day camp experience where youth in 3rd-6th grade participated in activities to learn about caves and to practice outdoor navigation skills using		Sixty nine individuals from Richland, Monroe, La Crosse, Vernon and Crawford counties participated in activities that were planned and led by		

Key Task: Structuring Data

- Via Codes and Document Variables
- Via Document Groups
- Via Document Colors

Structuring of Data via Codes

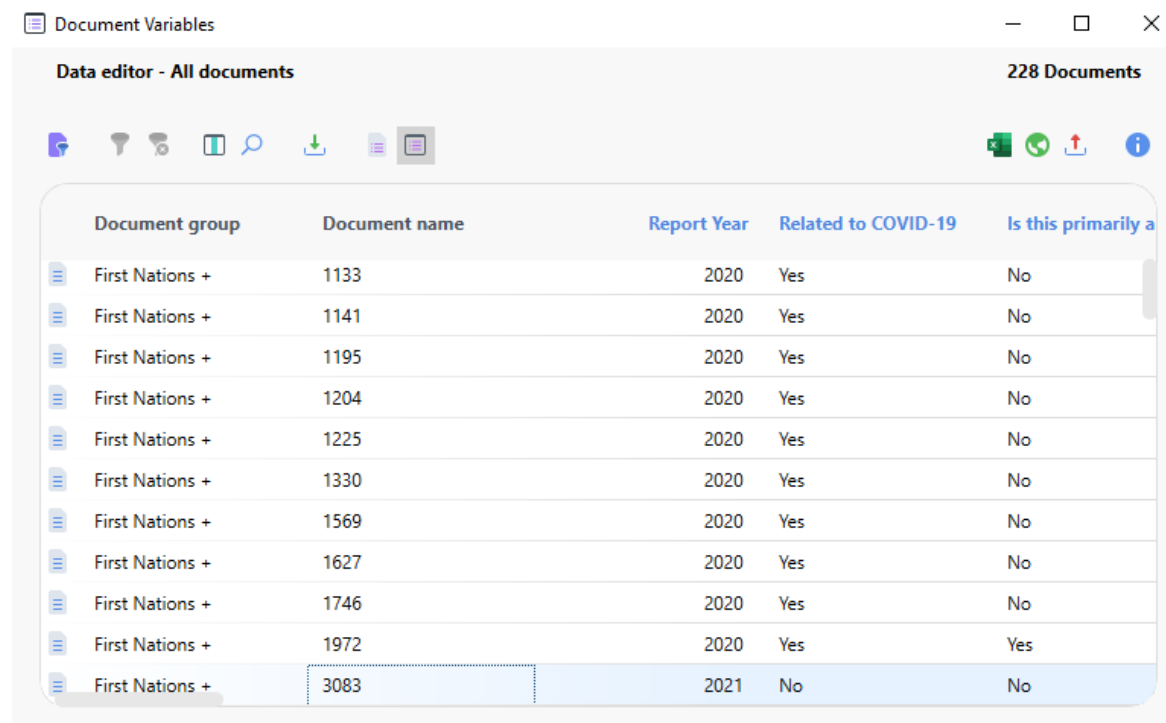
The screenshot displays the MAXQDA software interface, which is used for managing and analyzing large data sets. The interface is divided into several panels:

- Documents Panel (Top Left):** Shows a hierarchical view of documents. The 'Documents' folder contains 4258 items, and the 'First Nations +' sub-folder contains 234 items. A list of document numbers (1133, 1141, 1195, 1204, 1225, 1330, 1569, 1627, 1746) is shown with corresponding counts (23, 19, 18, 17, 16, 14, 18, 18, 10).
- Codes Panel (Bottom Left):** Shows a list of codes and their counts. The codes are: Home Office (228), Abstract Sentence (228), Outcome Statement/Supplemental I... (148), Report Year (228), Geography Served (228), and Institute (228).
- Document View (Right):** Displays the content of a selected document (4829, 11 Paragraphs). The document is titled 'Identified Outcome'. The content is structured into paragraphs, with the first paragraph starting with 'A series of presentations for Native Americans where they learn how cultural knowledge and water science can connect on water topics. This effort was designed to increase Native American confidence to be local advocates for water.' The second paragraph starts with 'Tribal Nations have been stewards of the local water resources in the area for a very long time. Cultural knowledge gained from this long relationship with the nature, also known as traditional ecological knowledge, is a needed perspective when addressing water issues. Roberts created and delivered a series of presentations, titled Izhi-Nibi: Water Ways, at the Lac Courte Orielles Ojibwe College St. Croix Outreach Center to community members. Topics included watersheds, manoomin (wild rice), and fish including ogaa (walleye) and name (lake sturgeon). The program was designed with input from the outreach center coordinator and an educator. The goal was to build connections between traditional ecological knowledge and western science. By building connections between cultural knowledge and science, Ojibwe community members will gain confidence in bringing their perspective to local water issues.' The third paragraph starts with '2021'.

Structuring via Document Variables

Document Variables

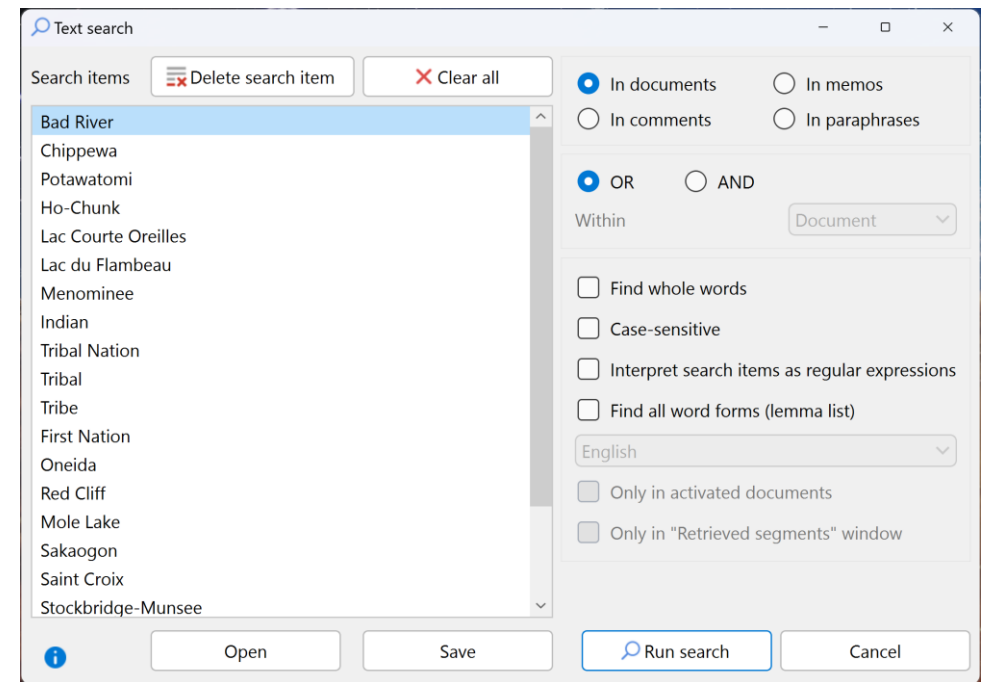
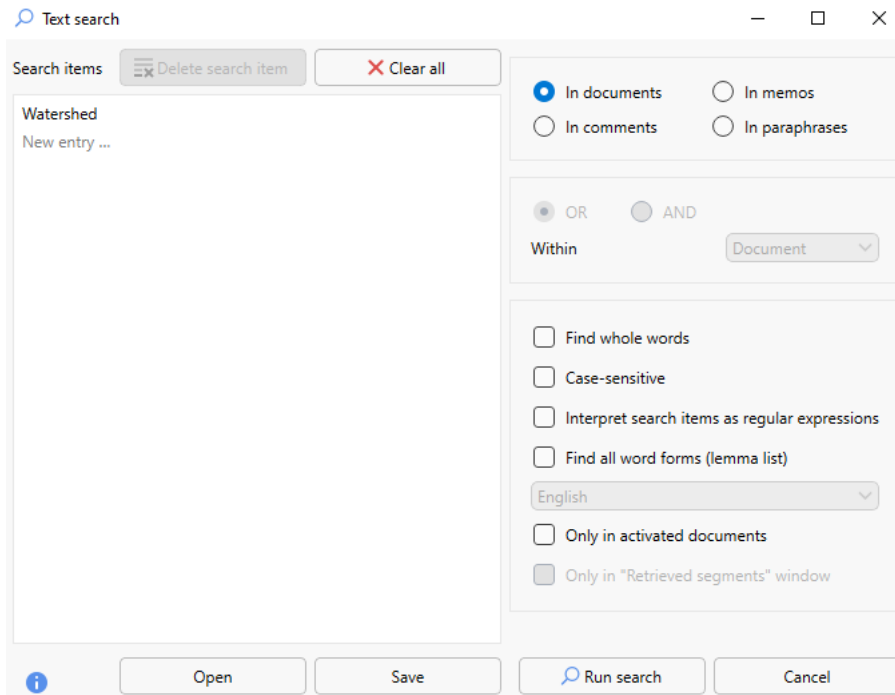
Data editor - All documents 228 Documents



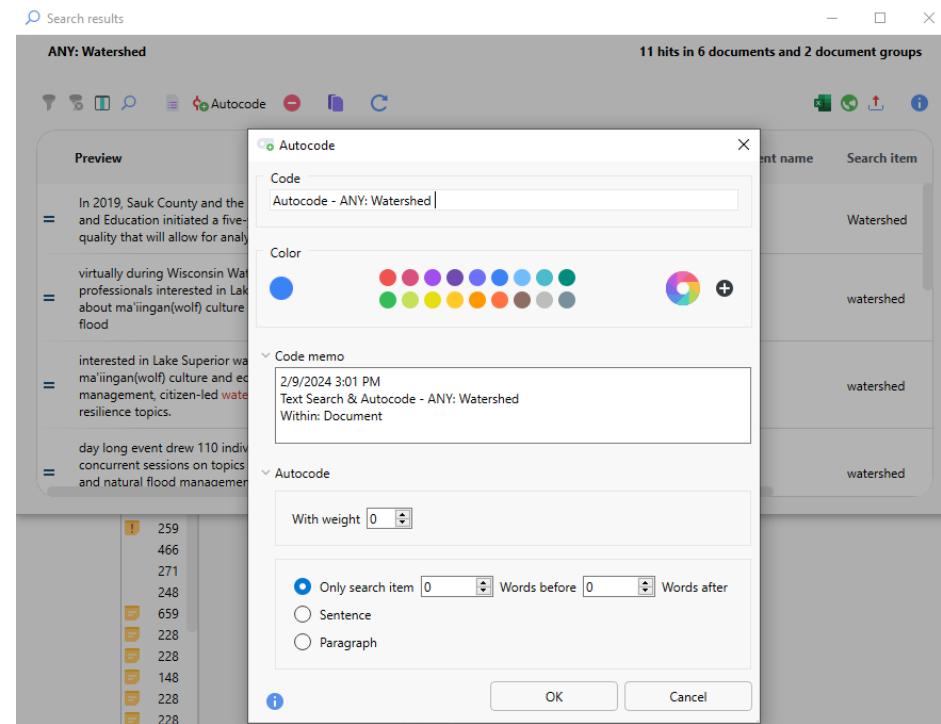
The screenshot shows the 'Document Variables' window in MAXQDA. It features a 'Data editor - All documents' tab with a toolbar containing icons for file operations, search, and data management. The main area displays a table with 5 columns: 'Document group', 'Document name', 'Report Year', 'Related to COVID-19', and 'Is this primarily a'. The table lists 11 documents, all from the 'First Nations +' group, with document IDs ranging from 1133 to 3083. The last row (ID 3083) is highlighted in blue.

Document group	Document name	Report Year	Related to COVID-19	Is this primarily a
First Nations +	1133	2020	Yes	No
First Nations +	1141	2020	Yes	No
First Nations +	1195	2020	Yes	No
First Nations +	1204	2020	Yes	No
First Nations +	1225	2020	Yes	No
First Nations +	1330	2020	Yes	No
First Nations +	1569	2020	Yes	No
First Nations +	1627	2020	Yes	No
First Nations +	1746	2020	Yes	No
First Nations +	1972	2020	Yes	Yes
First Nations +	3083	2021	No	No

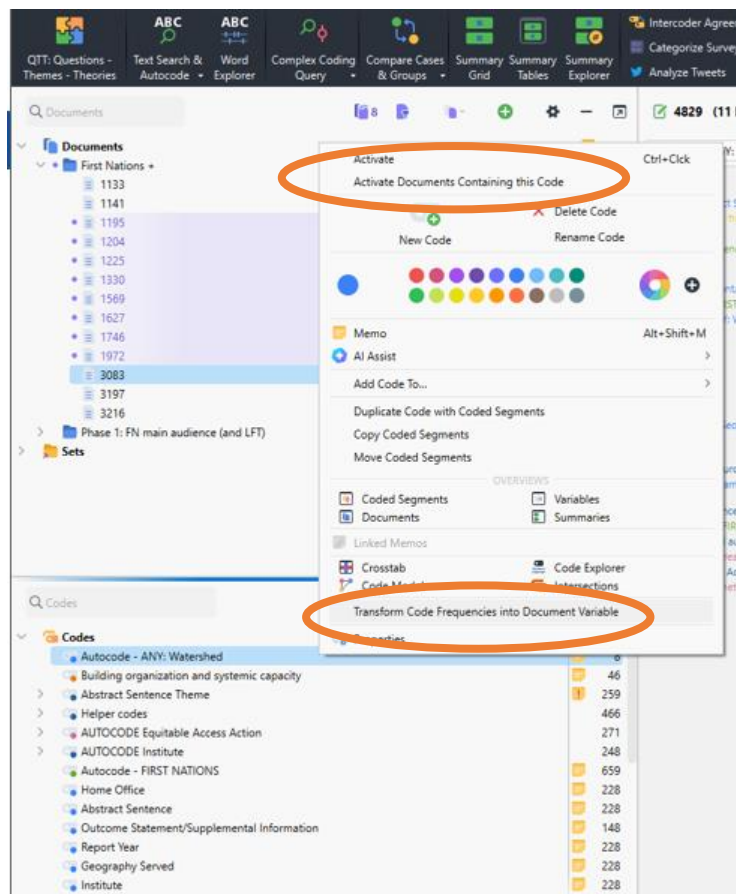
Importing unstructured Data? Use Auto-Coding



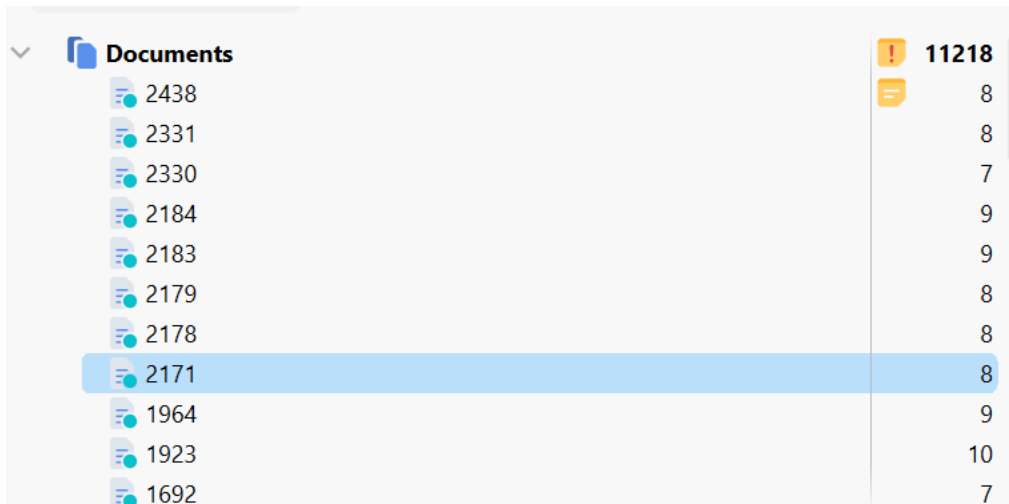
Importing unstructured Data? Use Auto-Coding!



No Document Variables? Transform Codes!



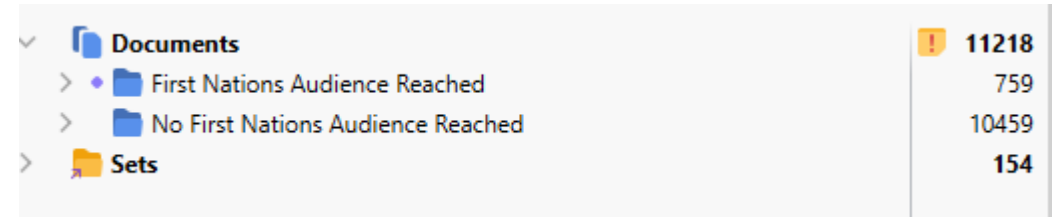
Structuring via Document Groups



A screenshot of a document management interface showing a list of documents under a 'Documents' folder. The list includes document IDs and a count of 11218. The document '2171' is highlighted in blue.

Documents	11218
2438	8
2331	8
2330	7
2184	9
2183	9
2179	8
2178	8
2171	8
1964	9
1923	10
1692	7

before

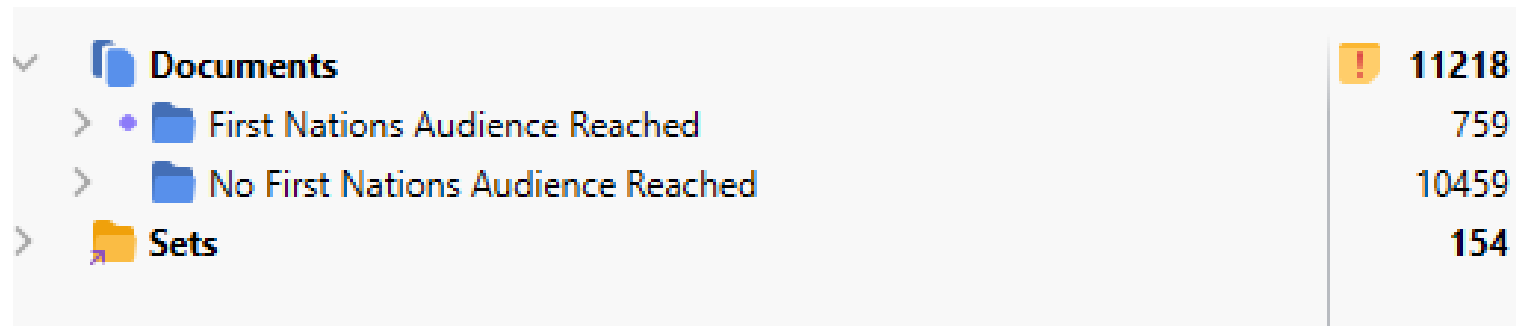


A screenshot of the same document management interface after grouping. The documents are now organized into two folders: 'First Nations Audience Reached' (759) and 'No First Nations Audience Reached' (10459). A 'Sets' folder is also visible with a count of 154. The total count remains 11218.

Documents	11218
> First Nations Audience Reached	759
> No First Nations Audience Reached	10459
> Sets	154

after

Structuring via Document Groups

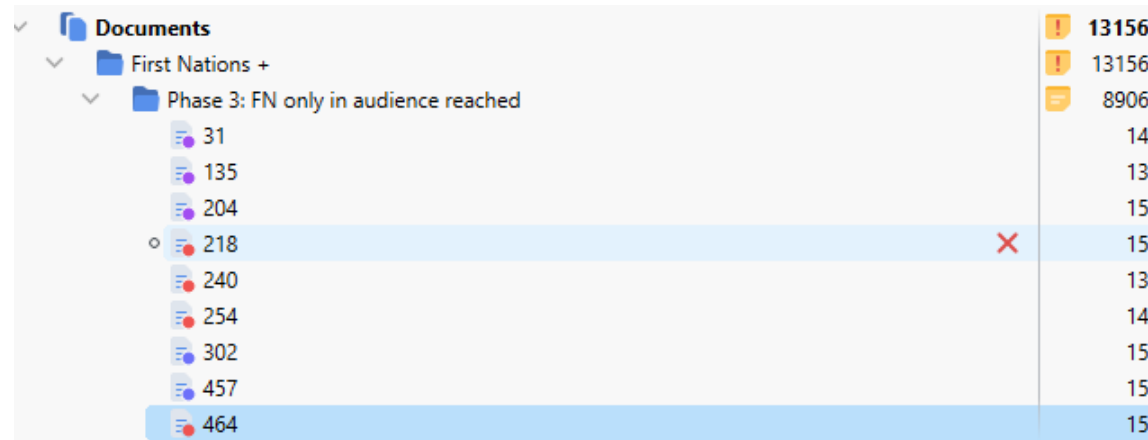


A screenshot of a document management interface. It shows a tree view on the left with a 'Documents' folder expanded, revealing two sub-folders: 'First Nations Audience Reached' and 'No First Nations Audience Reached'. Below these is a 'Sets' folder. On the right, a summary table shows the count of documents for each level: 'Documents' has 11218 (with a warning icon), 'First Nations Audience Reached' has 759, 'No First Nations Audience Reached' has 10459, and 'Sets' has 154.

Documents	11218
First Nations Audience Reached	759
No First Nations Audience Reached	10459
Sets	154

- Allows you to bulk move and bulk delete, and bulk re-import
- Sub-groups allow for more fine-grained structuring
- Helps manage distributed analysis in teams

Structuring via Document Colors



The screenshot shows a document management interface with a hierarchical tree structure. The tree is organized as follows:

- Documents
 - First Nations +
 - Phase 3: FN only in audience reached
 - 31
 - 135
 - 204
 - 218 (highlighted with a red 'X' in the status column)
 - 240
 - 254
 - 302
 - 457
 - 464

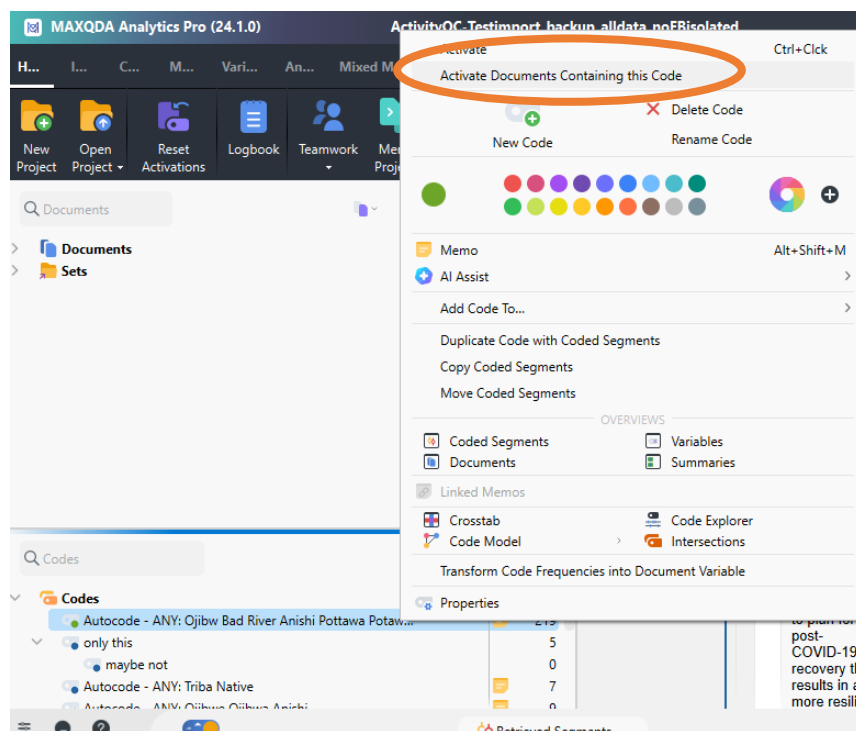
To the right of the tree, there is a column showing document counts and a status column. The counts are: 13156, 13156, 8906, 14, 13, 15, 15, 13, 14, 15, 15, 15. The status column contains a red 'X' next to the document 218.

- Allows you to bulk move
- Helps manage distributed analysis in teams

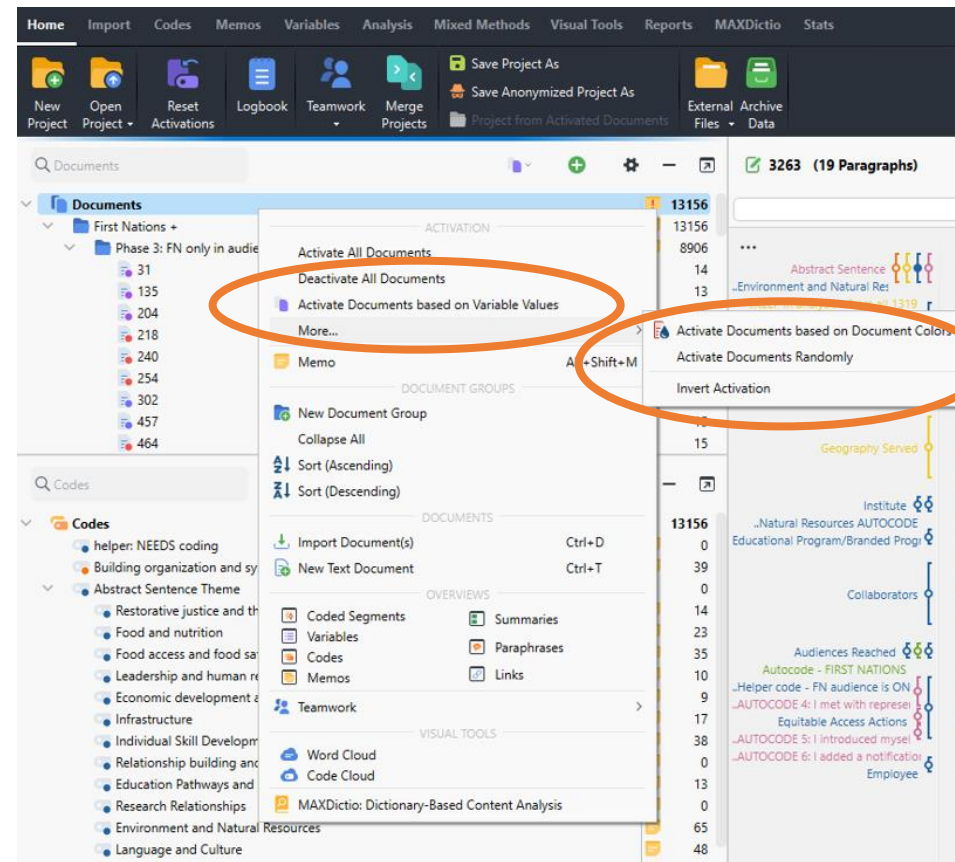
Key task: Bulk Movement & Deletion of Data

- Via Document Groups
- Via Text Search & Auto Code

Option 1: Activate documents by Code



Option 2: Activate Documents by Document Color or Variable Value



Option 3: Activate by Complex Coding Query

Complex Coding Query

Function
Intersection (Set)
Retrieve only the part segments that have at least X of the codes listed in "A" assigned to them (the number X is specified in "C").

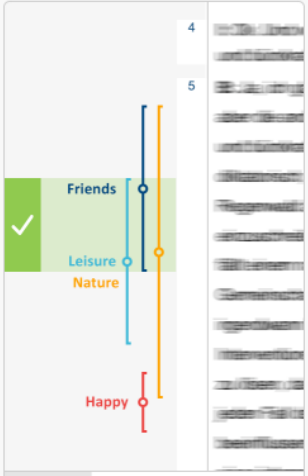
A: Codes
All activated codes Remove
Language and Culture
Food access and food safety
Leadership and human resource manage...
Drag and drop codes here

Options
☐ Only activated documents [3]
☐ Include subcodes
☐ Use weight filter
min: 0 max: 100
☐ Only coded segments of users:
cschmieder
allison.rogers

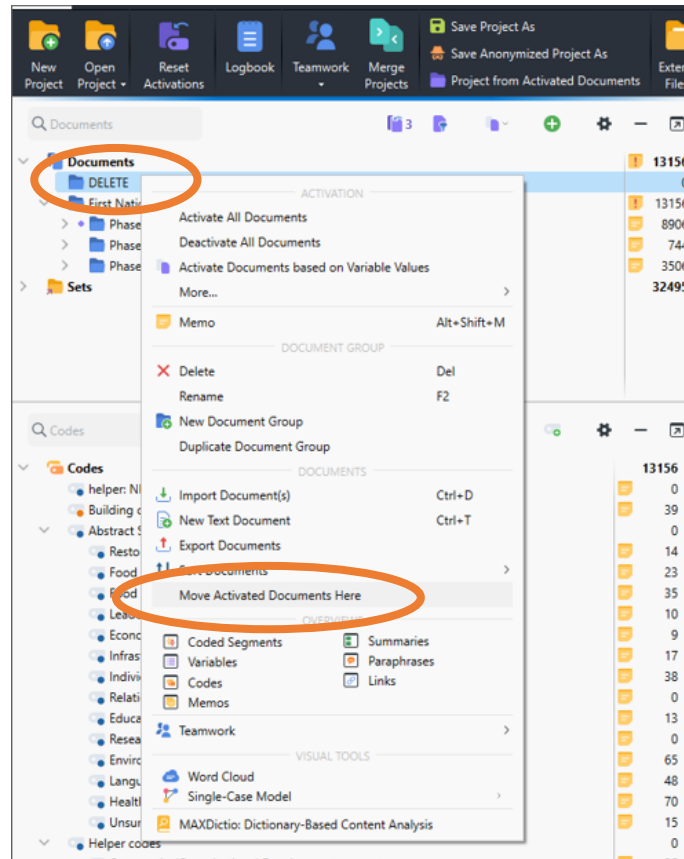
C: Min. number of codes
2 Codes

Segments found: 5

A: Codes Friends Happy Leisure
C: 2

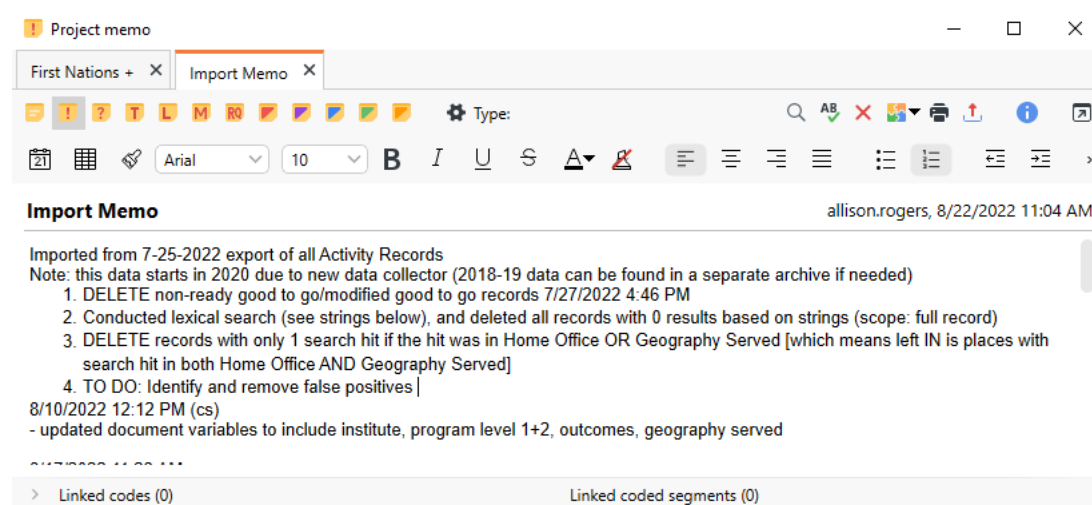


Move into new Document Groups & Delete via Document Group



PSA: Make Backups!

Key Task: Documentation



Time for Questions!