



# QualiAnon – Das Qualiservice Anonymisierungstool für Textdaten

Dr. Kati Mozygamba

Workshop der NFDI-Sektion Ethical Legal & Social Aspects (ELSA)

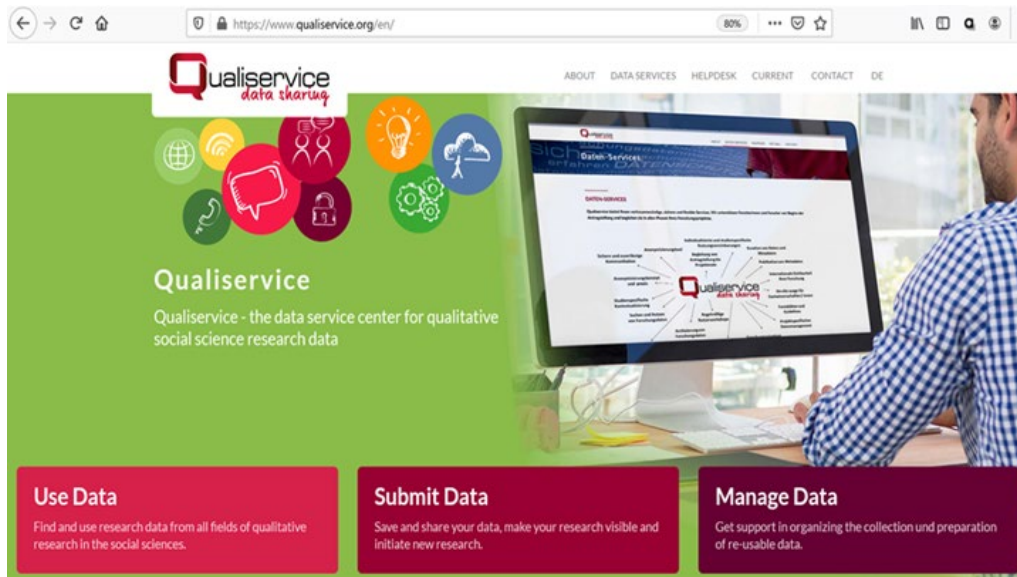
12.-13. Oktober 2022, Berlin

# Agenda

1. Background
2. Funktionalitäten von QualiAnon
3. Zugang, Manual etc.



# Das Forschungsdatenzentrum Qualiservice



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

- FDZ für qualitative Forschungsdaten der Sozialforschung (themenungebunden)
- Z.B.: Interviewtranskripte und -aufnahmen, Beobachtungsprotokolle Feldnotizen, Fotos, Audiovisuelle Daten
- Konsortium: PANGAEA, SuUB, GESIS Leibniz Institut für Sozialwissenschaften, FID Sozial- und Kulturanthropologie
- Verortet an der Universität Bremen,
- Leitung: Prof. Dr. Betina Hollstein
- Akkreditiert durch den RatSWD, Mitglied von KonsortSWD

# Besonderheiten qualitativer Forschung & qualitativer Daten

- Beispiel: biographische Erzählung über Krankheitsbewältigung
- Beinhalten oft Methodenkombinationen für Erhebung und Analyse
- Offener und iterativer Forschungsprozess, Forschungsdesign wird im Prozess angepasst
- Vertrauensverhältnis zwischen Forschenden und Beforschten
- Große Vielfalt und Dichte des Materials
- beinhaltet i.d.R. Umgang mit personenbezogenen Informationen
- hohe Kontextdichte und Verweise innerhalb der Daten
- Detaillierte Kontextinformationen als Rahmeninformationen für Datenanalyse

>> hohe datenschutzrechtliche Anforderungen und forschungsethische Standards

# Herausforderungen für die Nachnutzung qualitativer Daten

- Zwei potenziell konfligierende Interessen: Persönlichkeitsrechte schützen und Forschung ermöglichen
- Anonymisierung und Pseudonymisierung als zwei der zentralen Instrumente
- Braucht Konzept, welches sozialwissenschaftlich relevante Informationen erhält ohne „falsche Fährten“ zu legen
- einfache Löschungen/ Ersetzungen könnten zu falschen Ergebnissen führen

# Anonymisierung/ Pseudonymisierung

## Das Qualiservice-Konzept der “Flexiblen Anonymisierung”

- Sensible personenbezogene Daten werden mit unterschiedlich stark abstrahierenden sozialwissenschaftlich relevanten Informationen ersetzt
- Teilstandardisierung durch Listennutzung möglich
- Ersetzungen können auf unterschiedlichen Abstraktionsebenen wieder geöffnet werden
- Für bestimmte Fragestellungen können so spezifischere Informationen in ‘weniger’ anonymisierten Datensätzen bereit gestellt werden (onsite)



<http://dx.doi.org/10.26092/elib/192>

*Kati Mozygemba et al.*

**Anonymisierung und Pseudonymisierung qualitativer textbasierter Forschungsdaten – eine Handreichung**

University of Bremen – Qualiservice Working Papers

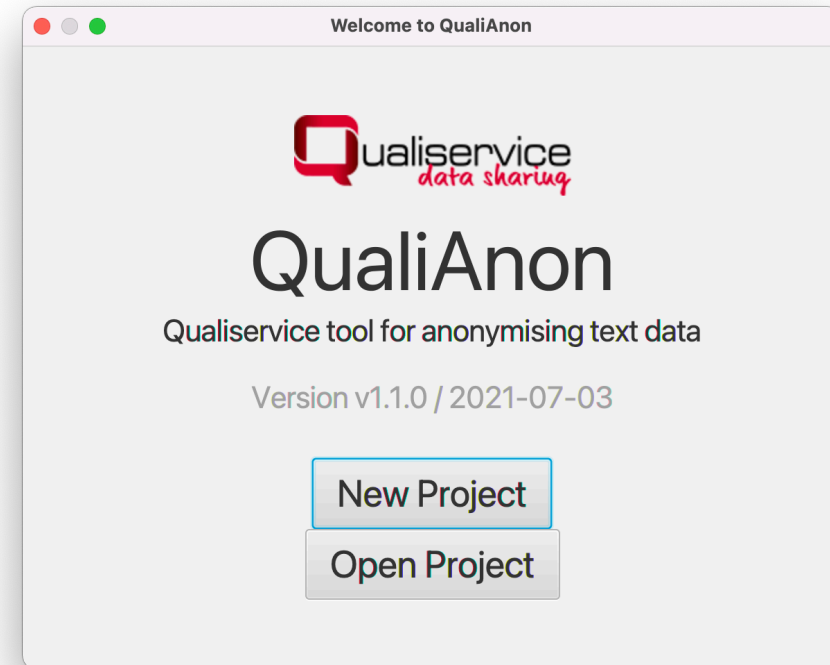
Im Erscheinen

# Beispiele

Kategorie	Subkategorie	Beispiel	Replacement			
			Original	Level 1	Level 2	Level 3
<b>Orte</b>	Land, Region, Bundesland, Stadt, Straße etc.	Berlin	Ort	Ort 1, Stadt in Deutschland	Ort 1, große Großstadt in Deutschland	Ort 1, große Großstadt in Deutschland mit mehr als 1 Mio. Einwohner
<b>berufliche Tätigkeiten</b>	Ausbildung, Berufe, berufliche Tätigkeiten	Altenpfleger	Beruf des Interviewten	Beruf des Interviewten, Dienstleistungsberufe und Verkäufer, ISCO-08: Code 05	Beruf des Interviewten, Betreuungsberufe, ISCO-08 Code 53	Beruf des Interviewten, Betreuungsberufe im Gesundheitswesen ISCO-08 Code 532
<b>Individuelle Charakteristik</b>	Krankheiten, Preise, Unfälle...	Schwerer Herzinfarkt	Krankheit	Krankheit 1	Schwere Erkrankung 1	Schwere Erkrankung des Kreislaufsystems

# QualiAnon – Tool für Textdaten

- Umsetzung des Konzeptes der „flexiblen Anonymisierung“
- Nutzung im FDZ (Kuration) und für Forschende
- User in control
- Nachhaltigkeit



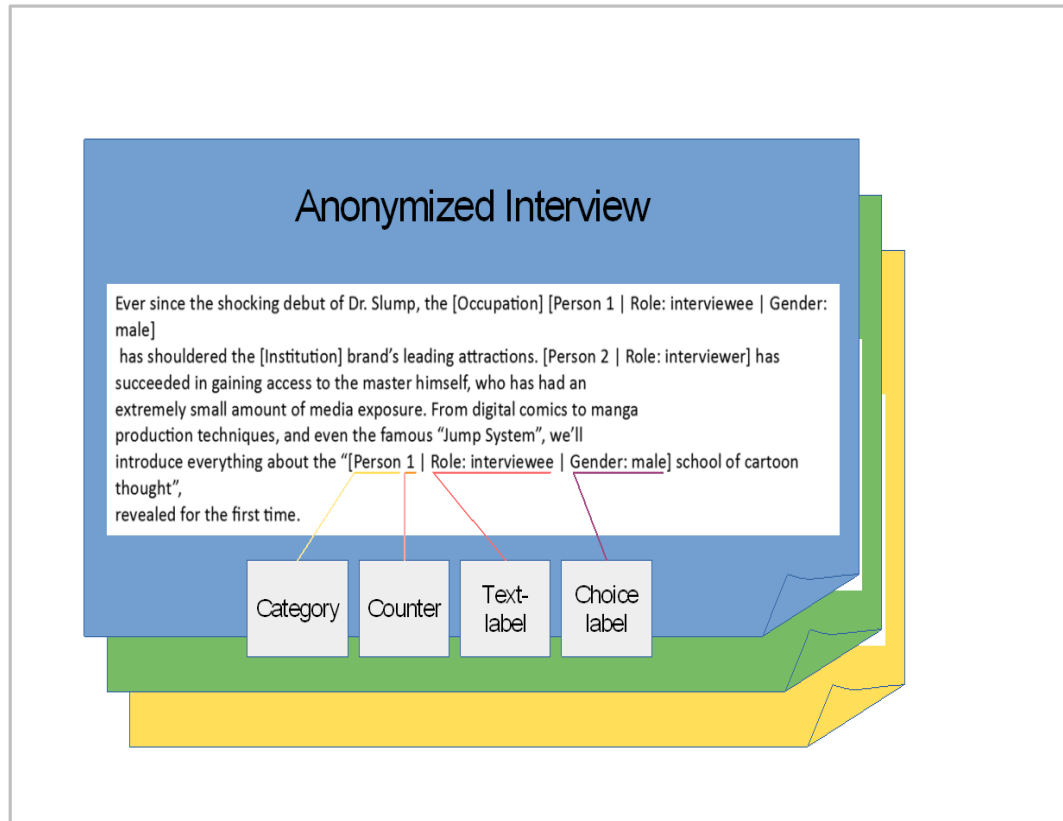
Reference: Tom Nicolai, Kati Mozygemba, Susanne Kretzer, Betina Hollstein (2021): QualiAnon – Qualiservice tool for anonymizing text data (version 1.1.0). Qualiservice. University of Bremen. Software available at: <https://github.com/pangaea-data-publisher/qualianon>



# Ausgewählte Funktionen

- halbautomatische Anonymisierung/ Pseudonymisierung von Textdaten (user in control)
- Ersetzungen auf Studien- und Fallebene
- Markieren & Ersetzen von sensiblen Informationen auf unterschiedlichen Abstraktionslevels
- Dokumentieren der Ersetzungen, schnelle Anpassungen ermöglichen
- Unterschiedlich stark abstrahierte Exporte
- Integration von (standardisierten) Listen
- Erstellen und Teilen von themenspezifischen Ersetzungsschemata
- Getrennte Speicherung

# Beispiel für abstrahierten Text



- Für jede Entität, die QualiAnon ersetzt, kann aus einem bearbeitbaren System von Ersetzungskategorien, eine Kategorie gewählt (z.B. Person, Beruf, Ort, etc.) werden.
- In eckigen Klammern findet sich die Information, die zur Ersetzung verwendet - hier: „[Person 1 | Role: Interviewee | Gender: male]“.
- Nummerierung von Ersetzungskategorien
- Flexible Anonymisierung ermöglicht unterschiedlich stark abstrahierte Exporte
- Export und Weiterentwicklung der Kategorienschemata

**Blackened interviews**

Ever since the shocking debut of Dr. Slump, the [M:947faa83-53d1-496a-868a-38032d7dc157:] [M:fd54a5a8-3af6-457a-9d36-37dadd9d5f5:] has shouldered the [M:2ca93753-0d37-4b17-a99c-d47f6065b3cb:] brand's leading attractions.

- One UUID for each occurrence -  
Every marked area has its own unique ID

**Marker file for each blackened interview**

Marker ID	Replacement ID	Original
947faa83-53d1-496a-868a-38032d7dc157	8fa05325-a72e-4a35-a8dc-f7fc2f54050c	cartoonist
fd54a5a8-3af6-457a-9d36-37dadd9d5f5	2ecb601e-c6bb-4afb-a588-96f6490f57e0	Akira Toriyama
2ca93753-0d37-4b17-a99c-d47f6065b3cb	a333dd59-c1a8-4fca-bd18-e6bcc66843ca	Jump

Connects the unique markers to common label data

**One replacements file for whole project**

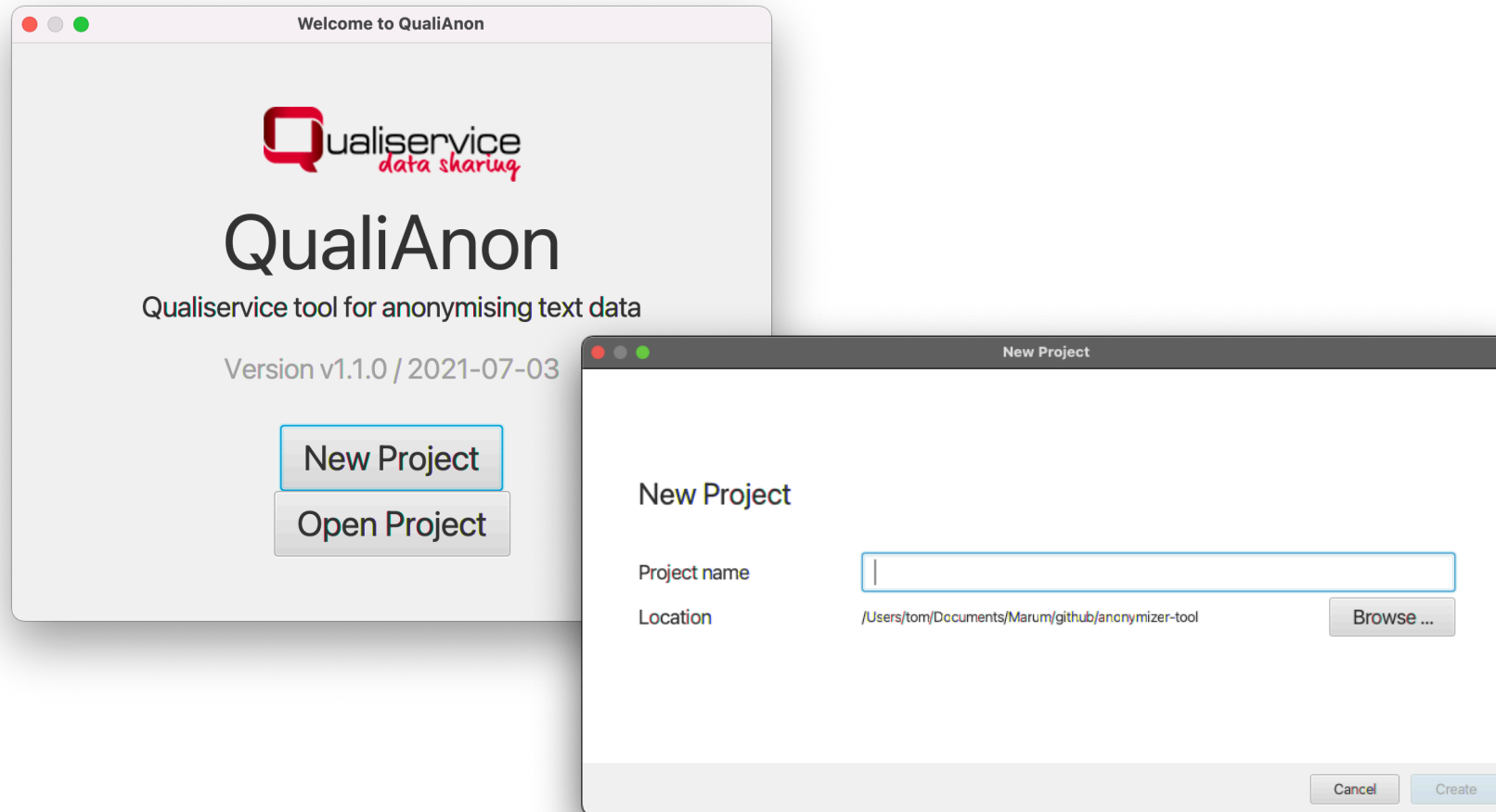
ID	Role	Gender
32b6e5ab-30a5-43ed-b438-fce42ffa2636	interviewer	
2ecb601e-c6bb-4afb-a588-96f6490f57e0	interviewee	male

- One UUID for each entity -  
Label data is used consistently throughout project interviews

Labels are defined in category schemes

- QualiAnon arbeitet mit einem System von Universally Unique Identifiers (UUIDs)
- Das anonymisierte Interview („blackened interview“) enthält die Marker-ID, die bei der Bearbeitung im Originaltext gesetzt wird.
- Die Replacement-ID bezieht sich auf die jeweils gewählte Übersetzung.
- In einem weiteren Dokument werden Marker-IDs und Replacement-IDs einer Studie sowie die Originalinformation gespeichert.

# Blick ins QualiAnon-Tool



File Edit View Anonymization Analysis Help

Categories

Location

Pseudonym:  C S Mark Mark All

Description:  List Lookup

Typ:  Apply

Bundesland:

Gemeinde:

▼ Replacements

Originals	Category	Pseudonym	Replacement	#
Katja Meier	Person	Person 1	Female	2
Zwickau	Location	Location 1	Größere Mittels...	4
SPIEGEL ONLINE	Person	Person 2		32

▼ Project: Mein Projekt

- Testinterview 1\_Gruene.txt
- Testinterview 2\_Politik.txt
- Testinterview 3\_Wirtschaft.txt
- Testinterview 4\_Krankheiten.txt
- Testinterview 5\_Russland.txt

Search: Meier Previous Next

(5) [Person 1], Jahrgang 1979, stammt aus [Typ=Größere Mittelstadt]. Studium der Politikwissenschaft, Soziologie in Jena, Münster und Estland. Seit 2015 für die Grünen im sächsischen Landtag, bei der Wahl 2019 Spitzenkandidatin. Seit Dezember 2019 ist Meier Justizministerin in der Kenia-Koalition unter CDU-Ministerpräsident Michael Kretschmer.

(6) Als es in der Silvesternacht in Leipzig zu Angriffen Linksextremer auf Polizisten kam, wurde die gewaltverherrlichende Liedstrophe aus Meiers Punkband im Internet verbreitet. Der stellvertretende AfD-Fraktionsvorsitzende im sächsischen Landtag, Sebastian Wippel, forderte den Rücktritt der Ministerin. Meier bezog daraufhin Stellung: Sie verurteile jede Form von Gewalt, das gelte auch für die Ausschreitungen in Leipzig.

(7) Im Interview mit dem SPIEGEL erklärt sie, was es mit dem Lied und ihrer Punkrockvergangenheit auf sich hat.

(8) SPIEGEL: Frau Meier, haben Sie damit gerechnet, dass diese alten Liedtexte irgendwann ausgegraben werden?

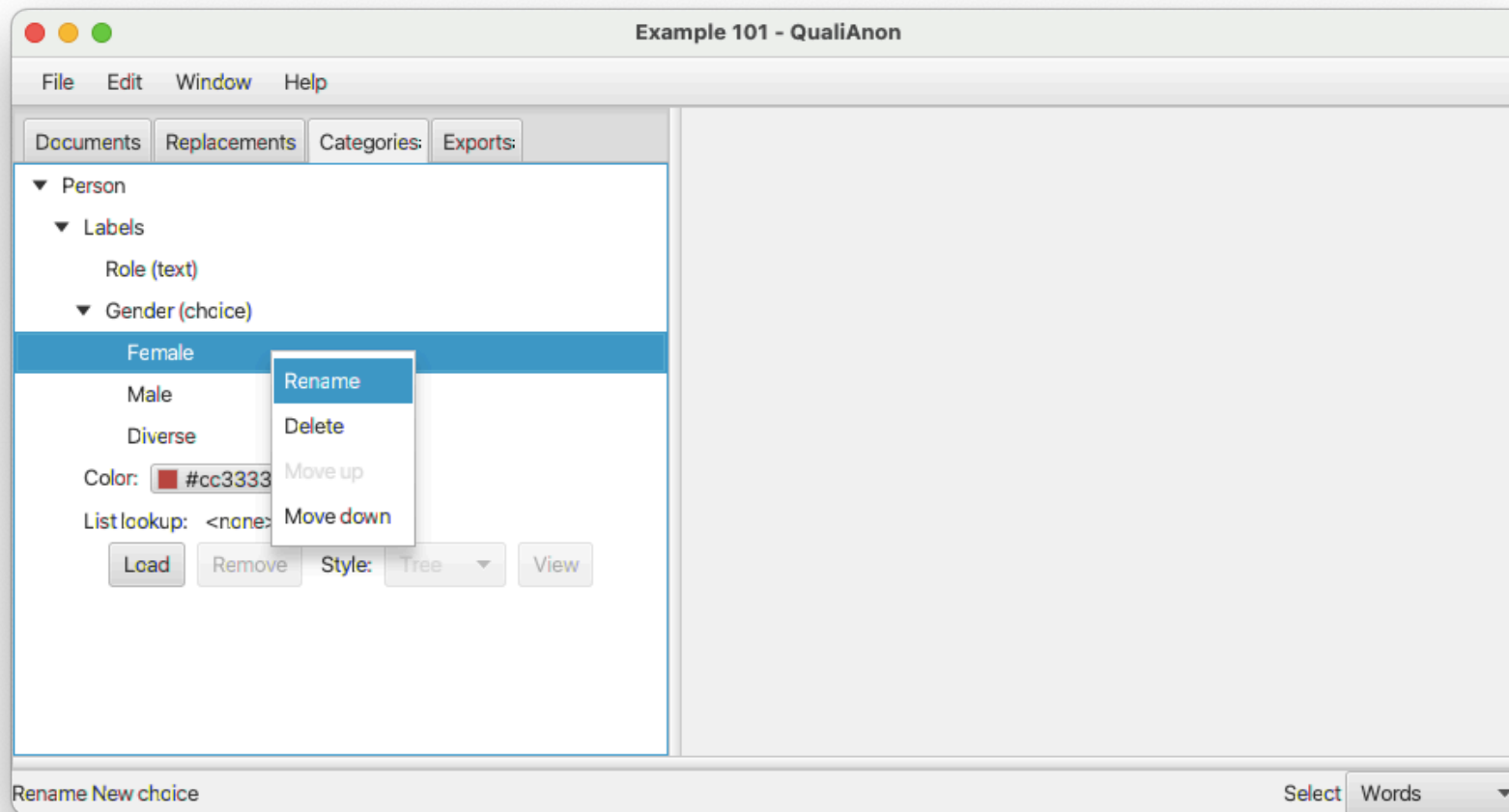
(9) Meier: Ich habe im Landtagswahlkampf immer erzählt, dass ich als Jugendliche in einer Punkband gespielt habe. Und dass da Lieder dabei sind, die ich heute nicht mehr spielen würde. Auch wenn ich die Texte selbst gar nicht geschrieben habe. Wir hatten nicht so viele Auftritte, kaum Publikum. Es ist absurd, dass das jetzt in einen Zusammenhang mit den Vorkommnissen in Leipzig gestellt wird.

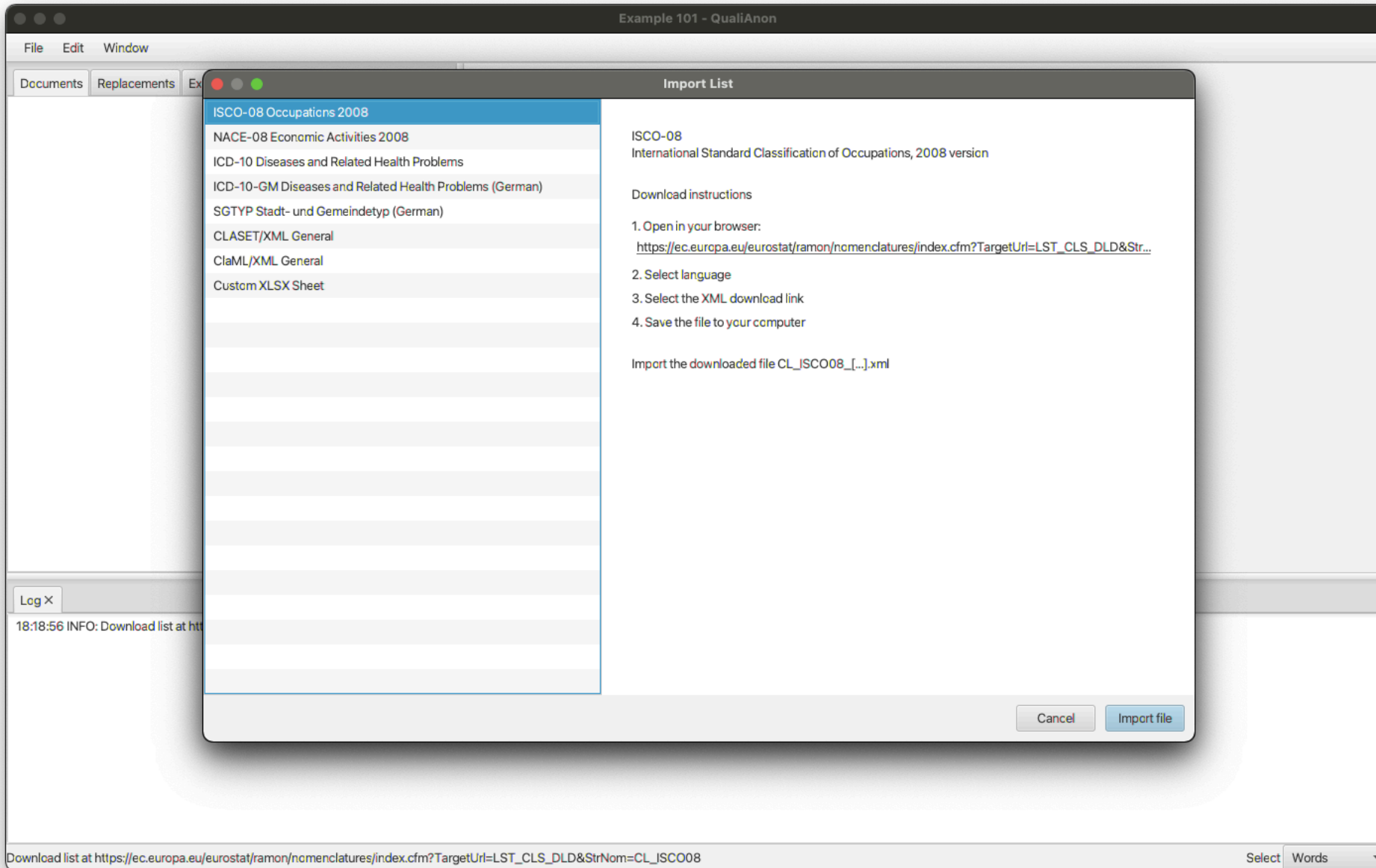
(10) SPIEGEL: Wo kam das her?

(11) Meier: Das weiß ich nicht. Das sind anonyme Accounts in den sozialen Netzwerken. Aber die Leute, die das auch auf Twitter weiterverbreitet haben, waren AfD-Landtagsabgeordnete, die sich klar und deutlich zu erkennen gegeben haben.

(12) SPIEGEL: "Advent, Advent - ein Bulle brennt". Warum haben Sie solche Lieder gesungen?

(13) Meier: Man muss meine Punkvergangenheit zeitlich einordnen. Das war in den Neunzigerjahren, ich war 15, 16 Jahre alt, als ich in der Band angefangen habe. Viele Jugendliche in Ostdeutschland hatten damals ähnliche Biografien. Arbeitslosigkeit der Eltern nach der Wende, ein starkes Aufkeimen von Rechtsextremismus, Fremdenfeindlichkeit. Nicht nur in Ostdeutschland natürlich, sondern auch im Westen, in





The screenshot shows a web application window titled "Example 101 - QualiAnon" with a menu bar (File, Edit, Window) and a toolbar (Documents, Replacements, Ex...). An "Import List" dialog box is open, displaying a list of categories on the left and details on the right. The "ISCO-08 Occupations 2008" category is selected. The right side of the dialog shows the name "ISCO-08 International Standard Classification of Occupations, 2008 version", download instructions, and a URL: [https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST\\_CLS\\_DLD&Str...](https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_CLS_DLD&Str...). The dialog has "Cancel" and "Import file" buttons at the bottom right. A log window at the bottom left shows the message "18:18:56 INFO: Download list at https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST\_CLS\_DLD&StrNom=CL\_JSCO08". The bottom right of the application window has a "Select" dropdown menu currently set to "Words".

Category	Description
ISCO-08 Occupations 2008	ISCO-08 International Standard Classification of Occupations, 2008 version
NACE-08 Economic Activities 2008	
ICD-10 Diseases and Related Health Problems	
ICD-10-GM Diseases and Related Health Problems (German)	
SGTYP Stadt- und Gemeindetyp (German)	
CLASET/XML General	
ClaML/XML General	
Custom XLSX Sheet	

Download instructions

1. Open in your browser:  
[https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST\\_CLS\\_DLD&Str...](https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_CLS_DLD&Str...)
2. Select language
3. Select the XML download link
4. Save the file to your computer

Import the downloaded file CL\_JSCO08\_[...].xml

Example 101 - QualiAnon

File Edit Window Help

Documents Replacements Exports

Testinterview\_1\_Gruene X

Quelle:  
<https://www.spiegel.de/politik/deutschland/katja-meier-saechsische-justizministerin-im-interview-ueber-punkvergangenheit-a-1303581.html>

Testinterview\_1\_Gruene

Testinterview\_2\_Politik

New Replacement for "Meier"

Category: Person

Pseudonym: Linda

Gender: Female

Cancel Insert

Log X

13:41:14 INFO: Loaded categories.xml  
 13:41:14 INFO: Loaded replacements file  
 13:41:14 INFO: Loaded Testinterview\_1\_Gruene.docx  
 13:41:14 INFO: Loaded Testinterview\_2\_Politik.docx

Loaded Testinterview\_2\_Politik.docx

Example 101 - QualiAnon

File Edit Window Help

Documents Replacements Exports

Category / Originals: Labels #

Kein Content in Tabelle

Testinterview\_1\_Gruene X

4 Die Grünen-Politikerin Katja Meier ist gerade erst in der  
 5 schwarz-grün-roten Landesregierung von Sachsen zur Justizministerin  
 6 ernannt worden, schon hat sie ihre rebellische Vergangenheit eingeholt.  
 7 Denn Meier spielte in jungen Jahren Bass bei einer Zwickauer Punkband,  
 8 den "Harlekins". In einem Lied skandiert der Sänger: "Advent, Advent -  
 9 ein Bulle brennt."  
 10  
 11 Zur Person  
 12 Katja Meier, Jahrgang 1979, stammt aus Zwickau. Studium der  
 13 Politikwissenschaften an der Universität Leipzig. Seit 2015  
 14 für die CDU in der sächsischen Landtagswahl 2019  
 15 Spitzenkandidatin für die CDU. Seit 2019 Justizministerin in der  
 16 Koalition mit der CDU. Vorgängerin: Michael Kretschmer.  
 17 Als es in der sächsischen Landesregierung um den  
 18 Polizisten Michael Kretschmer geht, offen Linksextremer auf  
 19 Punkmusik. Meier ist die Tochter des Liedstropheschreibers  
 20 AfD-Fraktion. Meier ist die Tochter des Liedstropheschreibers  
 21 forderte die CDU, Sebastian Wippel, log daraufhin Stellung:  
 22 Sie verurteilt die CDU, Sebastian Wippel, log daraufhin Stellung:  
 23 Ausschuss für die CDU, Sebastian Wippel, log daraufhin Stellung:  
 24 Im Interview mit dem SPIEGEL erklärt sie, was es mit dem Lied und ihrer  
 25 Punkrockvergangenheit auf sich hat.

Log X

13:41:14 INFO: Loaded categories.xml  
 13:41:14 INFO: Loaded replacements file  
 13:41:14 INFO: Loaded Testinterview\_1\_Gruene.docx  
 13:41:14 INFO: Loaded Testinterview\_2\_Politik.docx  
 13:43:50 INFO: Marked "Meier" with new replacement  
 13:46:37 INFO: Undo marked "Meier" with new replacement

Undo marked "Meier" with new replacement

Select Words



Example 101 - QualiAnon

File Edit Window Help

Documents Replacements Exports

▼ Export 101 \*

- ▼ Anonymization Settings
  - ▼ Category: Person \*
    - Original
    - Counting \*
    - Pseudonym \*
    - Gender \*
  - ▼ Category: Occupation \*
    - Original
    - Counting \*
    - Level 1
    - Level 2
    - Level 3
    - Level 4
    - Level 5
  - ▼ Category: Location \*
    - Original
    - Counting \*
    - Bundesland
    - Typ einfach
    - Typ differenziert
    - Gemeinde
  - ▼ Export Documents
    - Testinterview 1\_Gruene
    - Testinterview 2\_Politik

Testinterview 1\_Gruene [Export 101] ×

1 Quelle:  
2 <https://www.spiegel.de/politik/deutschland/katja-meier-saechsische-justizministerin-im-interview-ueber-punkvergangenheit-a-1303581.html>  
3  
4 Die Grünen-Politikerin Katja [Person 1 | Pseudonym: Linda | Gender: Female] ist gerade erst in der  
5 schwarz-grün-roten Landesregierung von Sachsen zur Justizministerin  
6 ernannt worden, schon hat sie ihre rebellische Vergangenheit eingeholt.  
7 Denn Meier spielte in jungen Jahren Bass bei einer Zwickauer Punkband,  
8 den "Harlekins". In einem Lied skandiert der Sänger: "Advent, Advent -  
9 ein Bulle brennt."  
10  
11 Zur Person  
12 [Person 1 | Pseudonym: Linda | Gender: Female], Jahrgang 1979, stammt aus Zwickau. Studium der  
13 Politikwissenschaft, Soziologie in Jena, Münster und Estland. Seit 2015  
14 für die Grünen im sächsischen Landtag, bei der Wahl 2019  
15 Spitzenkandidatin. Seit Dezember 2019 ist Meier Justizministerin in der  
16 Kenia-Koalition unter CDU-Ministerpräsident Michael Kretschmer.  
17 Als es in der Silvesternacht in Leipzig zu Angriffen Linksextremer auf  
18 Polizisten kam, wurde die gewaltverherrlichende Liedstrophe aus Meiers  
19 Punkband im Internet verbreitet. Der stellvertretende  
20 AfD-Fraktionsvorsitzende im sächsischen Landtag, Sebastian Wippel,  
21 forderte den Rücktritt der Ministerin. Meier bezog daraufhin Stellung:  
22 Sie verurteile jede Form von Gewalt, das gelte auch für die  
23 Ausschreitungen in Leipzig.  
24 Im Interview mit dem SPIEGEL erklärt sie, was es mit dem Lied und ihrer  
25 Punkrockvergangenheit auf sich hat.  
26  
27 SPIEGEL: Frau Meier, haben Sie damit gerechnet, dass diese alten  
28 Liedtexte irgendwann ausgegraben werden?  
29  
30 Meier: Ich habe im Landtagswahlkampf immer erzählt, dass ich als  
31 Jugendliche in einer Punkband gespielt habe. Und dass da Lieder dabei  
32 sind, die ich heute nicht mehr spielen würde. Auch wenn ich die Texte  
33 selbst gar nicht geschrieben habe. Wir hatten nicht so viele Auftritte,  
34 kaum Publikum. Es ist absurd, dass das jetzt in einen Zusammenhang mit  
35 den Verkommenen in Leipzig gestellt wird.

Disable Original Select Words

Example 101

Name	Änderungsdatum	Größe	Art
anonymized	Gestern um 23:05	--	Ordner
Testinterview 1_Gruene.docx	Gestern um 23:04	6 KB	Word-Dokument
Testinterview 2_Politik.docx	Gestern um 23:05	8 KB	Word-Dokument
categories	Gestern um 22:45	--	Ordner
categories.xml	Gestern um 22:45	1 KB	XML-Dokument
CL_ISCO08_20210210_160931.xml	Gestern um 16:09	169 KB	XML-Dokument
download-ref-sgtyp.xlsx	Gestern um 22:44	1,2 MB	3rd party formats
exports	Heute um 22:41	--	Ordner
Export 101	Heute um 17:10	--	Ordner
profile.xml	Heute um 17:10	1 KB	XML-Dokument
Testinterview 1_Gruene [export].docx	Heute um 17:10	6 KB	Word-Dokument
Testinterview 2_Politik [export].docx	Heute um 17:10	8 KB	Word-Dokument
identification	Gestern um 23:05	--	Ordner
replacements.xlsx	Gestern um 23:05	3 KB	3rd party formats
Testinterview 1_Gruene_markers.xlsx	Gestern um 23:04	3 KB	3rd party formats
Testinterview 2_Politik_markers.xlsx	Gestern um 23:05	3 KB	3rd party formats
trash	Gestern um 23:02	--	Ordner

# Wie bekomme ich Zugang zu QualiAnon?

- General information: [www.qualiservice.org](http://www.qualiservice.org)
- Access to software:  
E-mail to: [qualianon@uni-bremen.de](mailto:qualianon@uni-bremen.de)  
We will register you for the GitHub-project  
<https://github.com/pangaea-data-publisher/qualianon>
- Access to manual as google doc:  
[https://docs.google.com/document/d/1fLLYvsgXjh\\_p9p\\_E1fhikkIPQb19VUiltbRgaWQoD-M/edit](https://docs.google.com/document/d/1fLLYvsgXjh_p9p_E1fhikkIPQb19VUiltbRgaWQoD-M/edit)

