Virtual launch of the anonymization tool QualiAnon

Dr. Kati Mozygemba & Dr. Tom Nicolai

18. May 2021
Agenda

1. Background
2. Application scenarios
3. Functionalities of QualiAnon
4. Access to software, manual and further information
Why do we need an anonymization tool?

- RDC Qualiservice
- specialized on archiving and sharing qualitative research data in the social & behavioral sciences
- importance of data protection and research ethics
- anonymization and pseudonymization central instruments

www.qualiservice.org
Why do we need an anonymization tool?

- researchers need to deal with two potentially conflicting interests: protecting personal data vs. enabling research or re-use
- special contextual density of qualitative data
- anonymization and pseudonymization of qualitative data are comprehensive and time consuming

>> QualiAnon to support and facilitate the protection of sensitive information.
Scenario 1: Tool-Application in the context of Qualiservice

- allow for data protection and make research possible
- Anonymization is one part of the Qualiservice' concept of data protection
- „Flexible Anonymization“ (Kretzer 2013)
  - Replace sensitive information with information relevant for social science research
  - Document the replacements on different abstraction levels and keep them thus reversible
- A detailed guide on the anonymization of qualitative text data will be published soon.
Scenario 2: Tool-Application in the research project

- Researchers are confronted with questions of data security at different stages of their research process.
- Researcher need to provide anonymized data for different purpose.
- Need for flexible procedures.
- Qualiservice cooperates with researches, offers support, and checks the anonymization of a dataset. Data preparation is done in the research project.
Objectives of the tool development:

- QualiAnon should mark and replace sensitive information on different abstraction levels.
- It should document the replacements and allow for adaptations.
- It should allow for the integration of standardized lists (e.g., ISCO)
- It should allow to replace information on case basis as well as on study level.
- The export of versions differing in the amount of sensitivity should be possible.
- The user should be in control.
- The tool should be sustainable developed.
Ensuring sustainability

- QualiAnon is developed as an Open Source Tool.
- Software, manual and code will be accessible via GitHub (we will talk about this later again).
- GitHub allows for feedback and further development.
- There are no fees related to the use of QualiAnon.
How does QualiAnon work?
Goal No.1
Anonymization

Anonymized Interview

Ever since the shocking debut of Dr. Slump, the [Occupation] [Person 1 | Role: interviewee | Gender: male] has shouldered the [Institution] brand’s leading attractions. [Person 2 | Role: interviewer] has succeeded in gaining access to the master himself, who has had an extremely small amount of media exposure. From digital comics to manga production techniques, and even the famous “Jump System”, we’ll introduce everything about the ”[Person 1 | Role: interviewee | Gender: male] school of cartoon thought”, revealed for the first time.

- Additional conditions
  - Counting and labels must be consistent over all interviews of a project
  - Creation of different anonymizations must be fast and easy
Goal No.2
Separated Storage

- Additional conditions
  - Recombinable data
  - Standardized file formats

[Diagram showing storage and anonymization processes]
QualiAnon’s Data Model
Connect the UUIDs

Blackened interviews
Ever since the shocking debut of Dr. Slump, the [M:947faa83-53d1-496a-868a-38032d7dc157:] 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

<table>
<thead>
<tr>
<th>Marker ID</th>
<th>Replacement ID</th>
<th>Original</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>947faa83-53d1-496a-868a-38032d7dc157</td>
<td>8fa05325-a72e-4a35-a8dc-f7c2f54050c</td>
<td>cartoonist</td>
<td>interviewer</td>
</tr>
<tr>
<td>fd54a5a8-3af6-457a-9d36-37dadd9d5f5</td>
<td>2ecb601e-c6bb-4af8-a588-96f6490f57e9</td>
<td>Akira Toriyama</td>
<td>interviewee</td>
</tr>
<tr>
<td>2ca93753-0d37-4b17-a99c-d47f6065b3cb</td>
<td>a333d59-c1a8-4fca-bd18-e6bccc66843ca</td>
<td>Jump</td>
<td></td>
</tr>
</tbody>
</table>

Connects the unique markers to common label data

One replacements file for whole project

<table>
<thead>
<tr>
<th>ID</th>
<th>Role</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>32b6e5ab-30a5-43ed-b438-fce42f9a2636</td>
<td>interviewer</td>
<td></td>
</tr>
<tr>
<td>2ecb601e-c6bb-4af8-a588-96f6490f57e9</td>
<td>interviewee</td>
<td>male</td>
</tr>
</tbody>
</table>

Labels are defined in category schemes

- One UUID for each entity -
Label data is used consistently throughout project interviews
How can I get access?

• General information: [www.qualiservice.org](http://www.qualiservice.org)

• Access to software:
  E-mail to: qualianon@uni-bremen.de
  We will register you for the GitHub-project.

• Access to manual:
  E-Mail to: qualianon@uni-bremen.de

• We are looking forward to hear about your experiences.
Please see for further information

Web: [www.qualiservice.org](http://www.qualiservice.org)

Kontakt: [info@qualiservice.org](mailto:info@qualiservice.org)