A Practical Guide to Anonymisation Tools

The **anonymisation** of research data is a key step in enabling data sharing and reuse. As **data sharing** requirements are increasingly integrated into research funding and publication policies, researchers need **practical solutions** to prepare qualitative and quantitative data for dissemination while reducing re-identification risks.

A wide range of **tools** exists to support the anonymisation of qualitative and quantitative data. Please note that these tools cannot replace expert judgment. They provide valuable **support** for identifying and recoding personal data.

This brief guide provides a curated **selection** of commonly used tools in social sciences and highlights their key features to help you identify the solutions best suited to your research profile and dataset. This selection reflects the **landscape** at the time of writing. As these tools evolve rapidly, it is important to pay attention to software **versions** and **updates** when using them.

Key criteria for choosing a tool

 Data type	Is your source material qualitative or quantitative?
 Scale	What is the volume and sensitivity of the data?
 Technical skills	Do you have coding experience or require a graphical interface?
 Automation	Do you require manual control, semi-automation of fully automated detection?

Best practices for anonymising data, regardless of the tool

 Documentation	Document anonymisation decisions (e.g. maintain a data anonymisation log)
 Pilot testing	Pilot the anonymization process on a small sample and review results carefully
 Risk assessment	Assess contextual factors, sensitivity, and re-identification risks
 Secure environment	Work in a secure environment and avoid uploading non-anonymised data to cloud services

Qualitative anonymisation tools

Tools	Automation	Technical skills	Licence	Key points
UK Data Service Toolkit	Manual	Low	Open Access	+ Detailed guidelines + Compliance-oriented - Still largely manual - Limited compatibility with Mac OS
QualiAnon	Semi-auto	Low	Open Source	+ Suitable for long transcripts and field notes + Flexible anonymization options + UTF-8 encoding support - Tool still under development
Atlas.ti NVivo MAXQDA	Manual	Low	Proprietary	+ Visual interface integrated into QualAnalysis + Suitable for small projects - Not open source
FAMTAFOS	Semi-auto	Middle	Open Source	+ User-friendly desktop tool + NLP-based detection + Validated with motivated intruder test - Currently limited to English and Dutch
Microsoft Presidio	High	High	Open Source	+ Strong capability to detect identifiable information (names, dates, IDs) - Limited list of predefined personal identifiers

Quantitative anonymisation tools

Tools	Automation	Technical skills	Licence	Key points
sdcMicro	R or Shiny app	Medium – High	Open Source	+ Comprehensive statistical disclosure control (SDC) + Highly customizable for sensitive data + Provides quantitative risk metrics - Steeper learning curve
ARX	Graphical interface	Low	Open Source	+ Intuitive and easy to use + Strong visualization of risk and utility - No coding required - Less flexible for advanced custom anonymisation methods
AMNESIA	Graphical interface	Low	Open Source	+ Simple visual interface + Easy to use and flexible integration for developers - More limited anonymisation methods compared to sdcMicro

Resources

Additional resources on data management are available on the FORS website.

For anonymisation specifically, you may refer to our webinars Anonymisation of Qualitative Data and Anonymisation of Quantitative Data, which provide practical guidance for preparing data for sharing.

For more detailed methodological and strategic guidance, we also recommend consulting the following titles from the FORS Guides series:

- Data anonymisation (legal, ethical, and strategic considerations);
- Qualitative data anonymisation;
- Quantitative data anonymisation.

FORS
Data and services for the social sciences