

DE-IDENTIFYING / ANONYMISING YOUR DATA FOR SHARING

From the Global Collaboration on Traumatic Stress FAIR Data Workgroup

The basics

- Know your institutional / national regulations & requirements
- Build future de-identification / anonymisation into your data plan
- Direct identifiers
 - Remove (or don't collect if not needed)
 - Store separately from sensitive data
 - If / when feasible, delete master list connecting study IDs to direct identifiers
- Indirect / "quasi" identifiers
 - Which variables – alone or in combination – could be identifying?
 - Remove or generalize these – balance this with analytical value
 - Unusual values or combinations? Consider removing if particularly identifiable

Consider context for sharing

- How & where will data be shared?
- How open vs restricted?
- Ensure that users agree not to re-identify and to report any inadvertent re-identification



In traumatic stress research – consider:

- Public profile? Highly public, well-known event?
- Small or easily-identified group of people affected?
- Study data content - Potential for stigma or adverse consequences?

Want to systematically assess risk of re-identification?

- Learn about contemporary analytical approaches (k-anonymity & others)
- Use these assessments to inform further modifications to dataset, if needed

Learn about FAIR Data practices and tools: <https://www.global-psychotrauma.net/fair-tools>