Understanding De-identification Guidance and Practices for Research Data

Wentao Guo, Aditya Kishore, Paige Pepitone,[1] Adam Aviv,[2] Michelle Mazurek

University of Maryland, [1]NORC at the University of Chicago, [2]The George Washington University

✉ wguo5@umd.edu

@wentaochirps

Motivating examples

Academic researchers are studying restrictions on reproductive care.

They survey women in areas where **abortion is criminalized** about barriers to access.

Evaluators are contracted to assess the impact of foreign aid programs in conflict zones.

They survey residents about perceptions of **organized crime and terrorism**.

Public data has benefits

- Replication, meta-analysis
- Transparency for public funds
- Required by journals/funders

But de-id is challenging

- Traditional methods flawed
- Diff. privacy has accessibility & acceptability barriers

Our research

Analyzing guidance

Thematic analysis of 38 de-id guides (pub. post-2018)

- What techniques?
- Framing of outcomes
- Usability

Conducting interviews

Interviews of 26 experienced researchers and reviewers

- How do they de-id data?
- Perceptions of threats
- Challenges

Highlighted findings

Guides still skew towards traditional methods

- 36 out of 38 guides: **generalization** (coarsening)
- 28: **pseudonymization**
- 17: ***k*-anonymity**
- 11: **differential privacy**

Gaps in threat coverage

- Listing salary and medical diagnosis as non-identifying info
- Examples where deleted data can be deduced from context

Perceiving unlikely threats, practitioners use heuristic methods, fail to prevent singling out

"You could crosstab all variables in theory, but that would be like millions of crosstabs. It's not necessarily a scientific process. It's **more knowing what to look for**."

Funders & repositories sometimes push for weaker de-id

They felt if you've **removed all the really obvious** things—like name, state, town of residence, and date of birth—then that's probably enough.

SP2 SECURITY. PRIVACY. PEOPLE

GWU SEC

NSF

NORC

icons: flaticon.com