Accessing & Sharing Linked Data Within A Research Team

Tales from the field...

Prof Luke Sloan

SloanLS@Cardiff.ac.uk

@proflukesloan

Outline

- General Principles

- Data Sharing Agreements (DSAs)

- USIP Linked Survey & Twitter Data

- NatCen/Yonder Twitter Handles

- NatCen/Yonder Linked Survey & Twitter Data

- Lessons Learnt

General Principles

- Systematic processing

- Data reduction

- Controlled access

- Data deletion

- Negotiating ethical review (flexible, discursive)

- Contractual agreements (rigid, legal) e.g. legal basis for processing
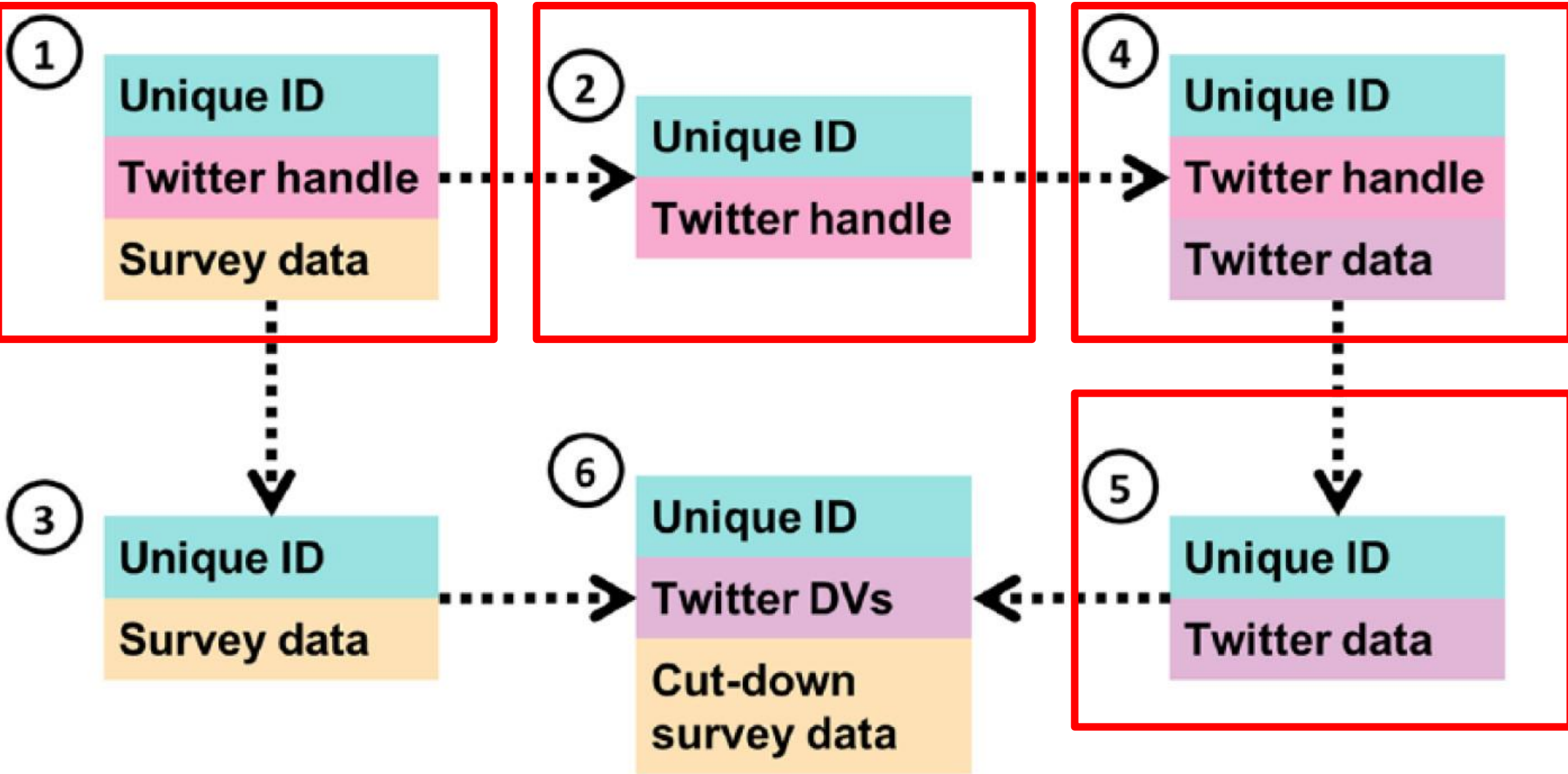
Data Sharing Agreements (DSAs)

- Need access to:
  - USIP data including linked Twitter and survey data
  - NatCen Panel/Yonder Twitter usernames
  - NatCen Panel/Yonder linked Twitter and survey data

- Separate agreements for all three

- Tripartite – too complicated, too many players!

USIP Linked Survey and Twitter Data

- Relatively simple

- Analysis of secondary data (collected outside of current project)

- RA signed up as ISER Research Associate (fulfills obligations)

- Access to secure servers via VPN

- Data deposited in shared folder

NatCen Panel/Yonder Twitter Handles

▶ Separate DSA required (no associate option)

▶ Data is disclosive with unique identifiers

▶ Researcher at Cardiff needs handles to access Twitter content

▶ How to ensure deletion of data at end of project
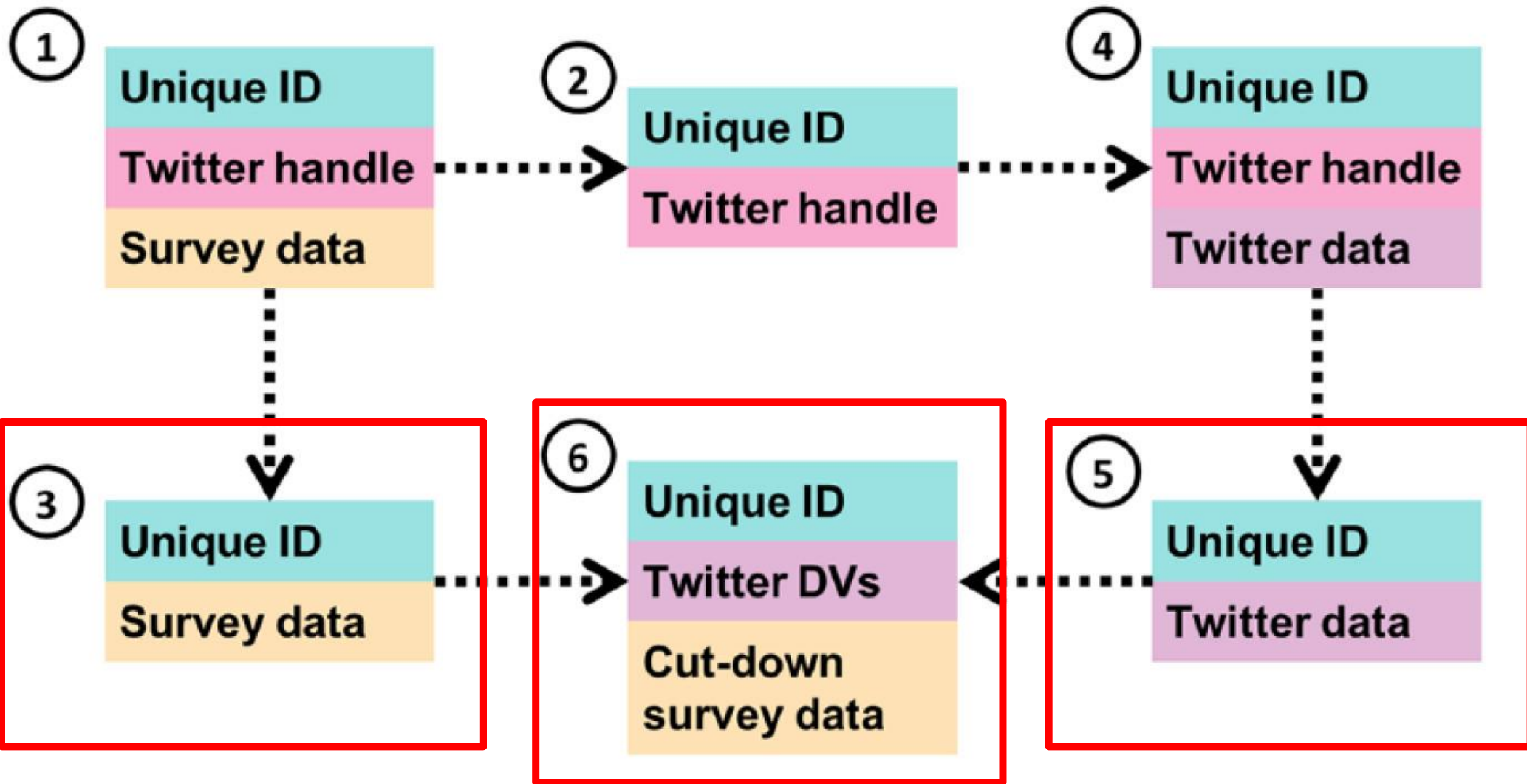
▶ Do you know how deletion works in the cloud?

Source: Sloan et al. (2020) https://journals.sagepub.com/doi/10.1177/1556264619853447

NatCen Panel/Yonder Twitter Handles

- Dedicated laptop provided

- Secure download/upload from/to NatCen servers

- Data only stored locally (no copies, nothing on the cloud)

- Laptop to be secure-wiped at end of project

- Possible only to identify that respondent is part of study

NatCen Panel/Yonder Linked Survey & Twitter Data

- Separate DSA required (on top of DSA for Twitter handles)

- Linked data is much more complicated

- Personal identifiers removed, but still disclosive

- Tweet data could be used to identify a survey respondent

① Unique ID / Twitter handle / Survey data

② Unique ID / Twitter handle

④ Unique ID / Twitter handle / Twitter data

③ Unique ID / Survey data

⑥ Unique ID / Twitter DVs / Cut-down survey data

⑤ Unique ID / Twitter data

Source: Sloan et al. (2020) https://journals.sagepub.com/doi/10.1177/1556264619853447

NatCen Panel/Yonder Linked Survey & Twitter Data

▶ Data deemed to be much higher risk

▶ 'Associate' role not an option – consequences?

▶ Accordingly, agreement is with Cardiff University, not the individual

▶ Dedicated, secure laptop provided by NatCen

▶ Secure laptop accesses servers via VPN

Lessons Learnt

▶ Erasing from cloud-based services is not simple

▶ Such a high level of control involves compromise on backups

▶ Availability of software on secure devices

▶ Bilateral agreements impact on three-way collaboration

▶ Continuing access post-project?