

an interaction model for
de-identification of **human data**
held by **external custodians**

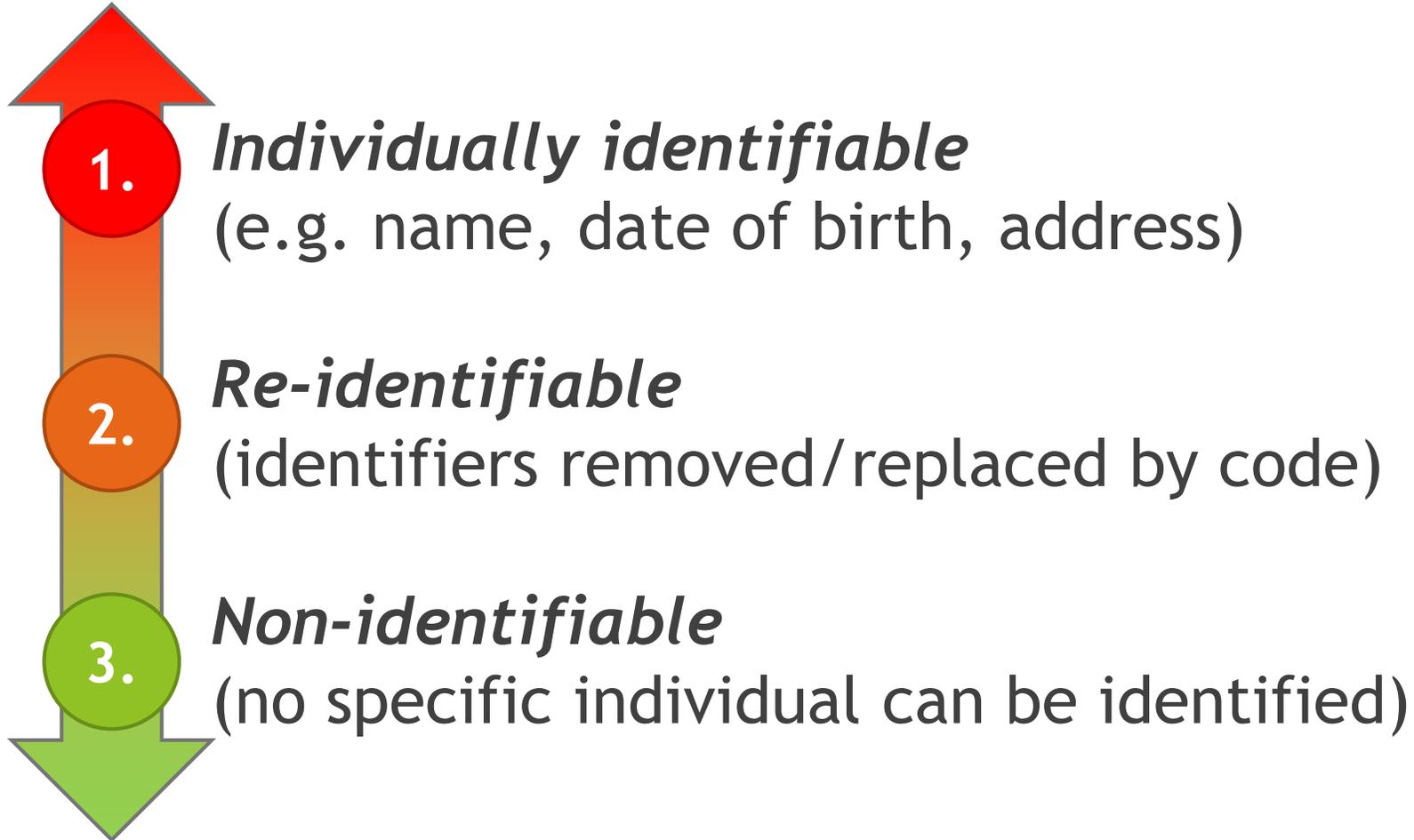
Andrew Simmons, Maheswaree Kissoon Curumsing, Rajesh Vasa

Applied AI Institute, Deakin University

Data Identifiability

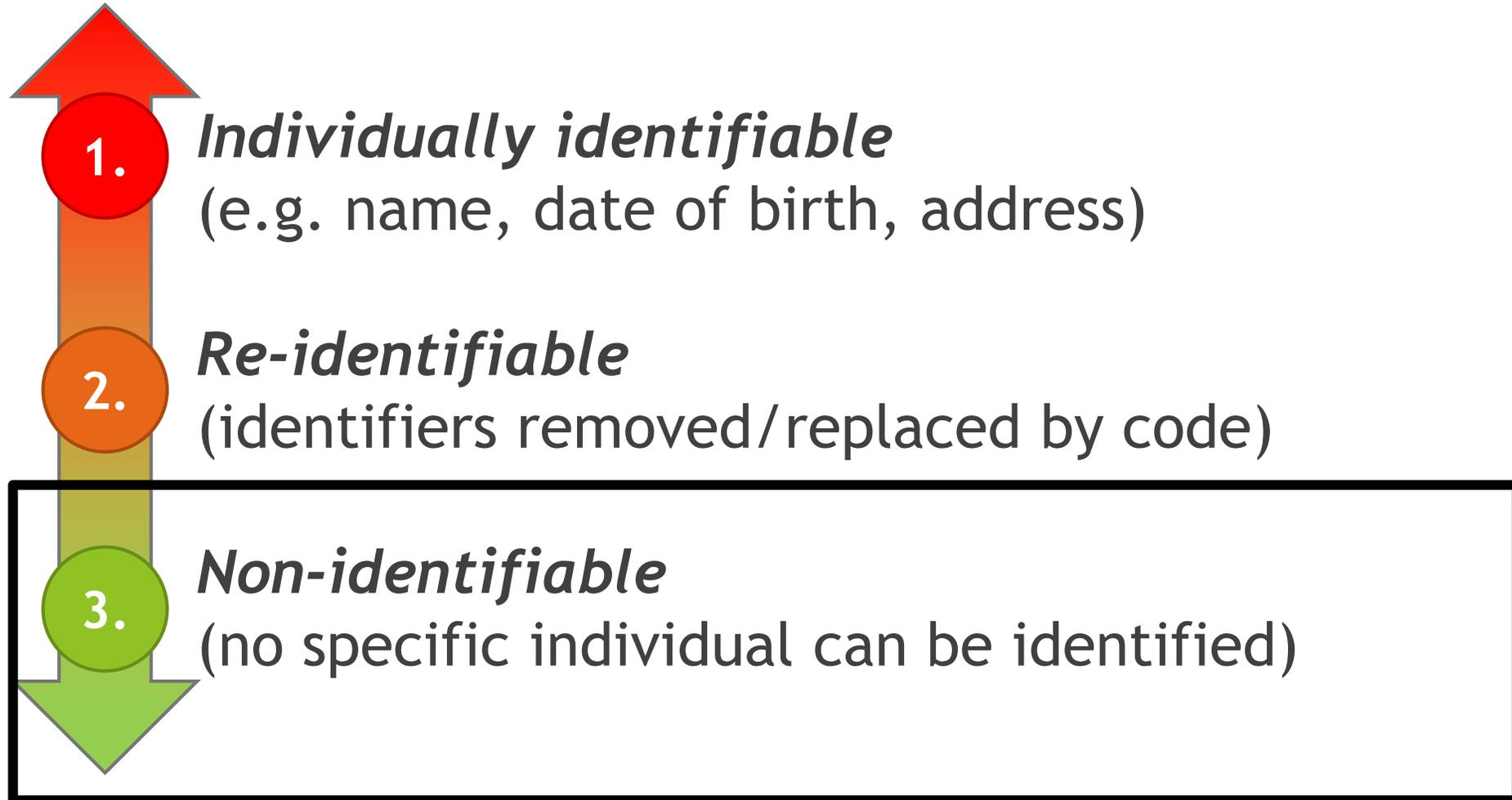
-  ***Individually identifiable***
(e.g. name, date of birth, address)
-  ***Re-identifiable***
(identifiers removed/replaced by code)
-  ***Non-identifiable***
(no specific individual can be identified)

Data Identifiability



*National Statement on Ethical Conduct in Human Research, 2007
(updated 2018)*

Data Identifiability



For ethics exemption, only pre-existing **non-identifiable** data can be used.

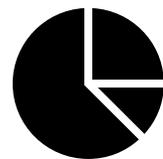
Existing Approach

Data custodian de-identifies data



A. **Too little** detail:

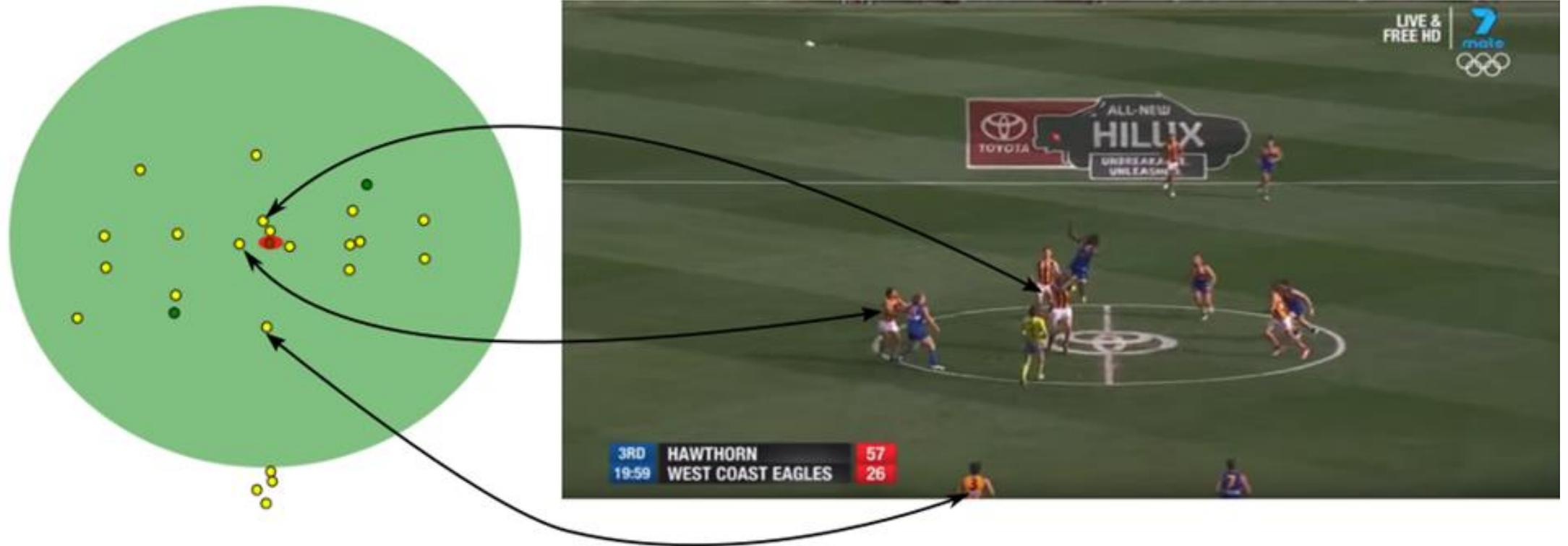
- ▶ Data insufficient to answer interesting research questions



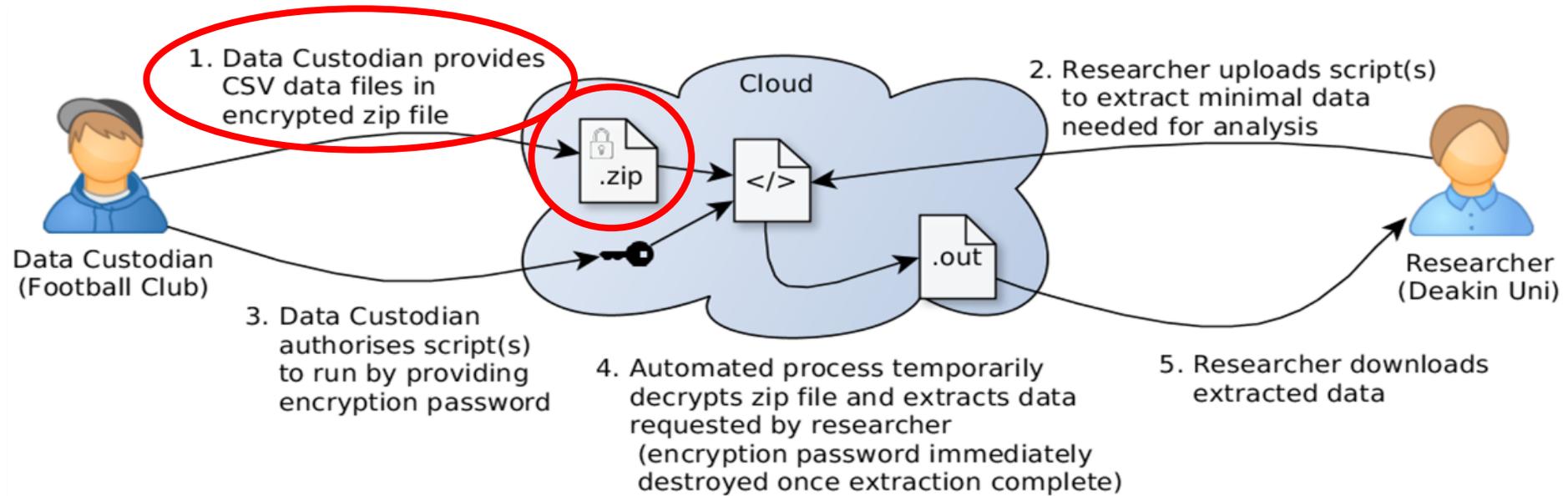
B. **Too much** detail:

- ▶ Researcher in violation of human ethics protocol
- ▶ Legal issues for data custodian
- ▶ Privacy risk to data subjects

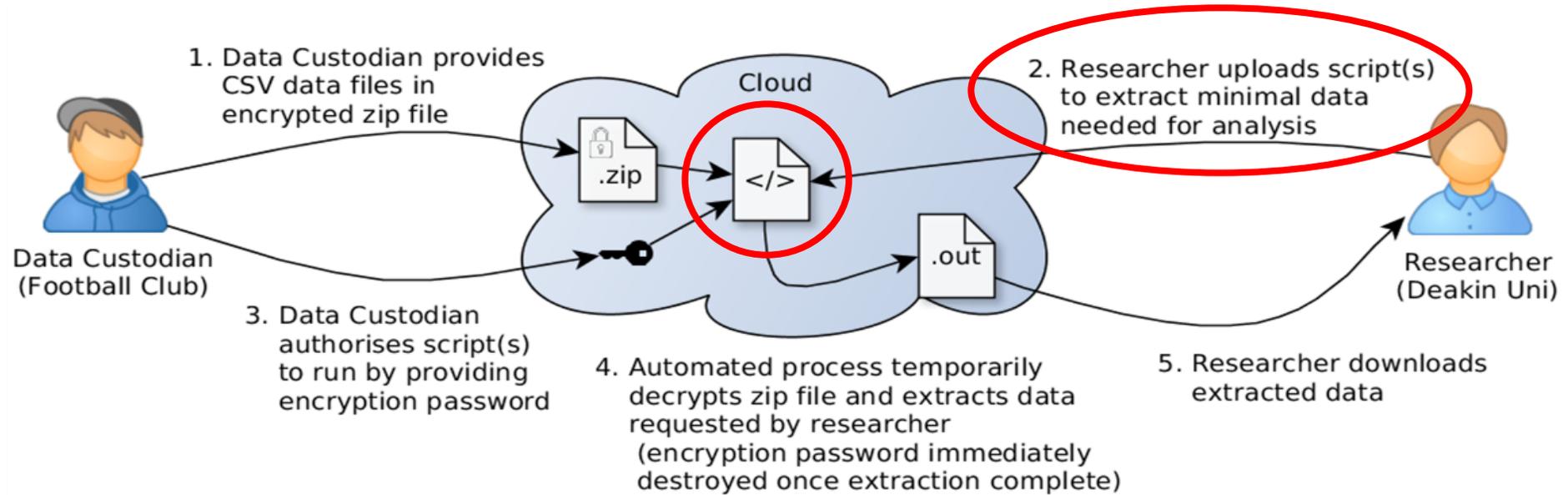
Example



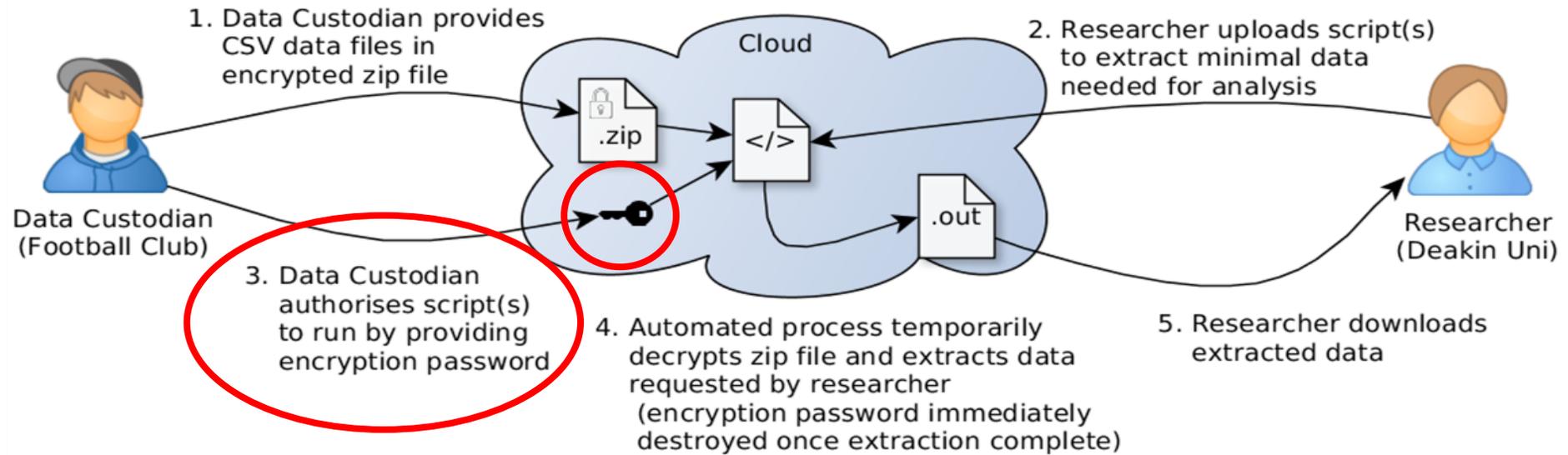
Solution



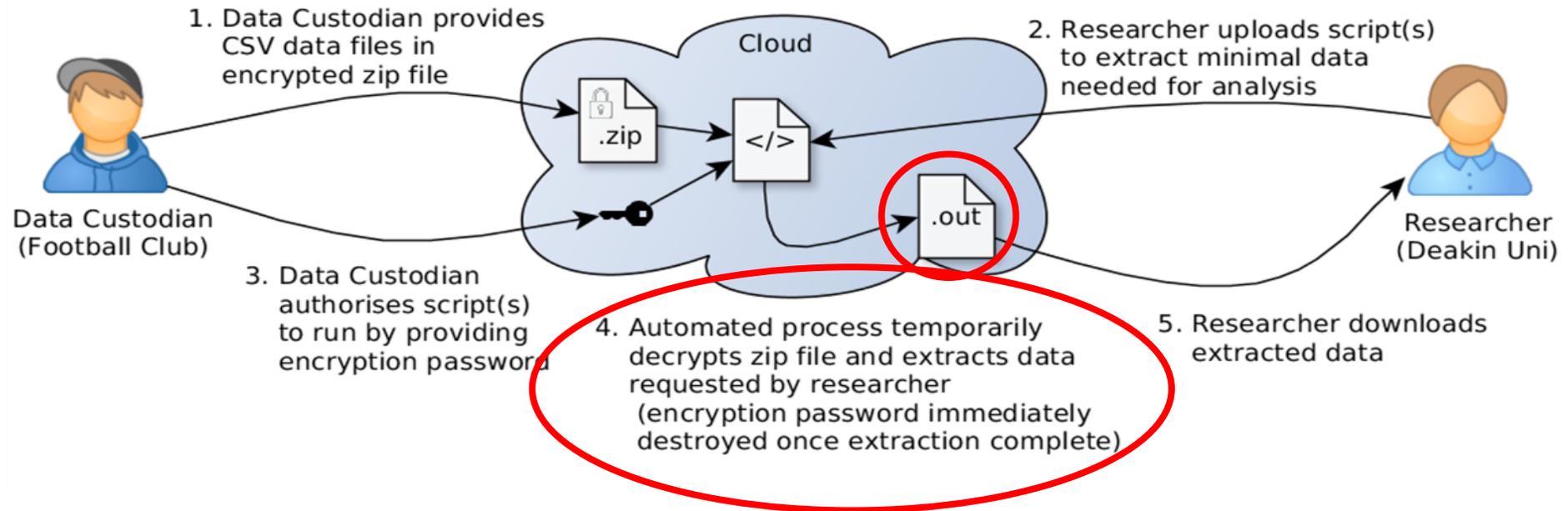
Solution



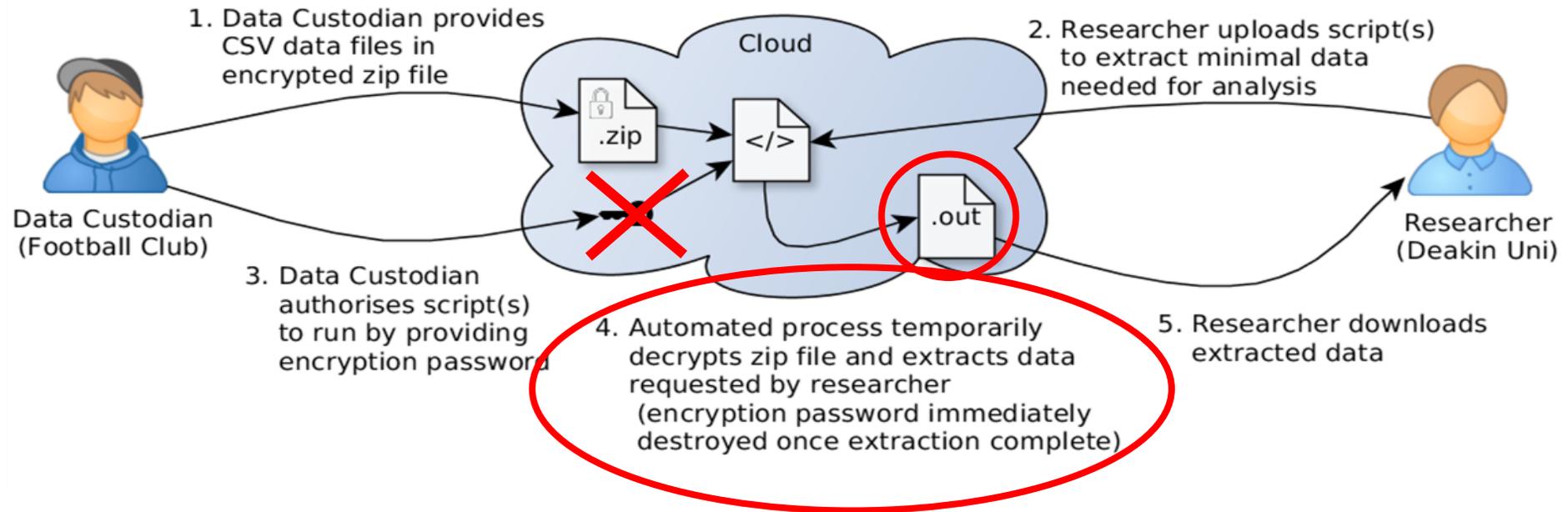
Solution



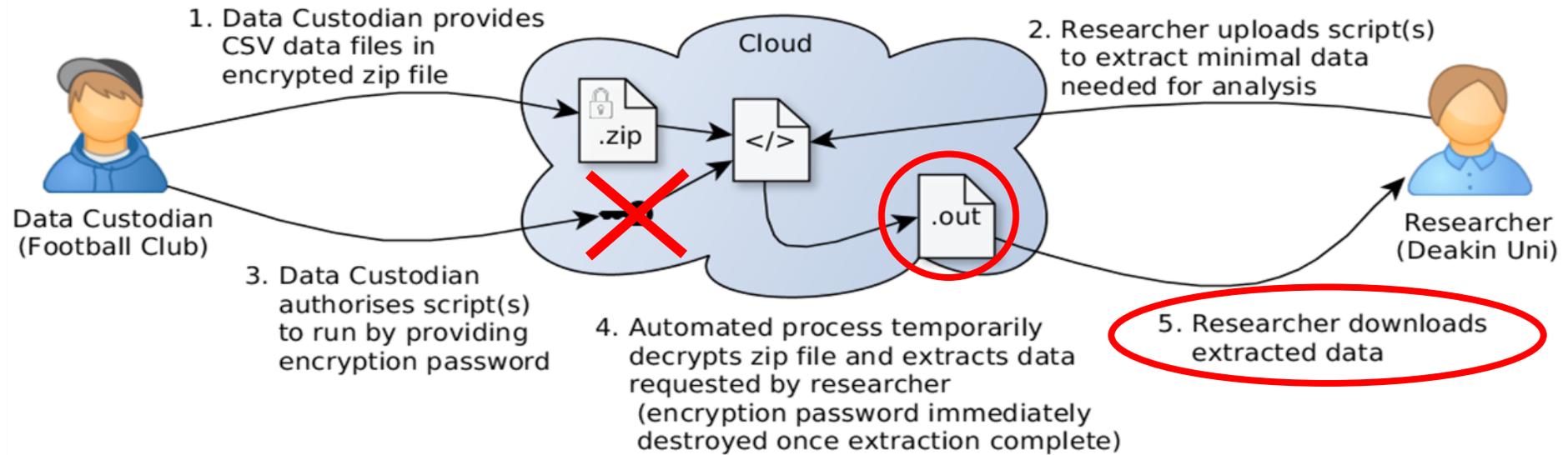
Solution



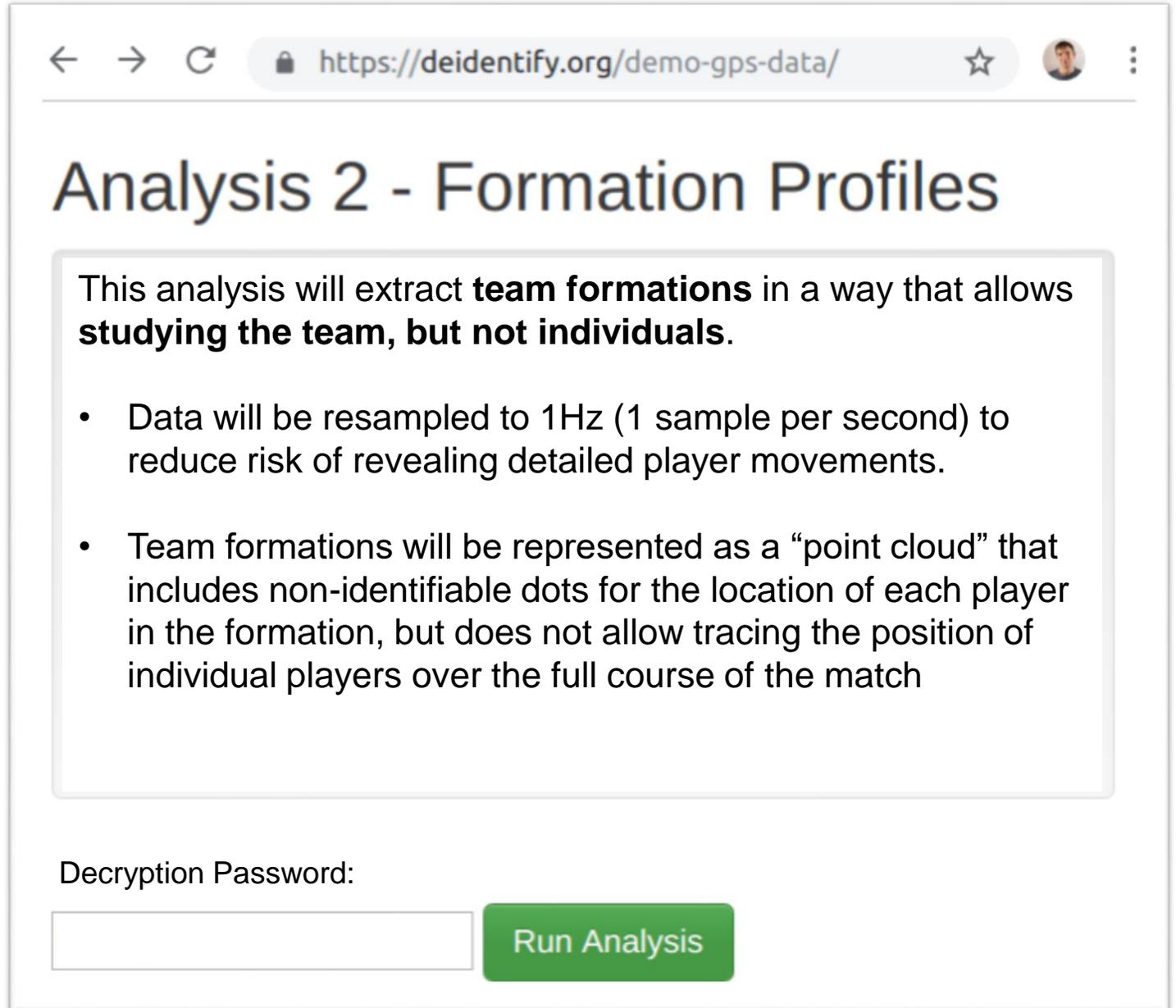
Solution



Solution



Deidentify.org



The screenshot shows a web browser window with the address bar displaying <https://deidentify.org/demo-gps-data/>. The page title is "Analysis 2 - Formation Profiles". Below the title, a text box explains the analysis process: "This analysis will extract **team formations** in a way that allows **studying the team, but not individuals**." Two bullet points follow: "Data will be resampled to 1Hz (1 sample per second) to reduce risk of revealing detailed player movements." and "Team formations will be represented as a 'point cloud' that includes non-identifiable dots for the location of each player in the formation, but does not allow tracing the position of individual players over the full course of the match". At the bottom, there is a "Decryption Password:" label, an empty text input field, and a green "Run Analysis" button.

← → ↻ <https://deidentify.org/demo-gps-data/> ☆  ⋮

Analysis 2 - Formation Profiles

This analysis will extract **team formations** in a way that allows **studying the team, but not individuals**.

- Data will be resampled to 1Hz (1 sample per second) to reduce risk of revealing detailed player movements.
- Team formations will be represented as a “point cloud” that includes non-identifiable dots for the location of each player in the formation, but does not allow tracing the position of individual players over the full course of the match

Decryption Password: