Data Linkage
Identifying records that correspond to the same entity in multiple data sources. Many application areas:
- Health Sector
- National Censuses
- National Security
- Digital Libraries
- E-Commerce
- Social Science Research
- Economics

Genealogy Linkage
Using historical records to reconstruct family histories. Motivation:
- For individuals: desire to learn about ancestors
- For historians and historical demographers: answering research questions.
- Change in family sizes over time
- Relation of wealth and causes of death
- Patterns in occupation changes
- And many more…

Available Records
- Most common
  - Vital events (Birth, Marriage, Death)
  - Censuses
- Less common
  - Adoption, Religious, Medical
  - Immigration, Naturalisation
  - Occupational, Educational
- Errors
  - Missing data, typos
  - Spelling variations
  - Nicknames instead of full names
  - Transcription errors
  - Translation / transliteration errors

Automatic Linkage
- Doing linkage by hand is time consuming, tedious, and error-prone.

Computer Science can help!
- Automatic linkage algorithms
- Start by preparing the data: digitising, transcribing, cleaning
- Robust comparison methods
- Using statistical tools
- Efficient algorithms that are able to receive input from domain experts
- Can handle very large amounts of data
- We are happy to talk more about the underlying techniques!

Contact us
- ozgur.akgun@st-andrews.ac.uk
- School of Computer Science
- University of St Andrews

The family tree of Ludwig Herzog von Württemberg (ruled 1568-1593) Painting from 1585 - Currently in Landesmuseum Stuttgart