

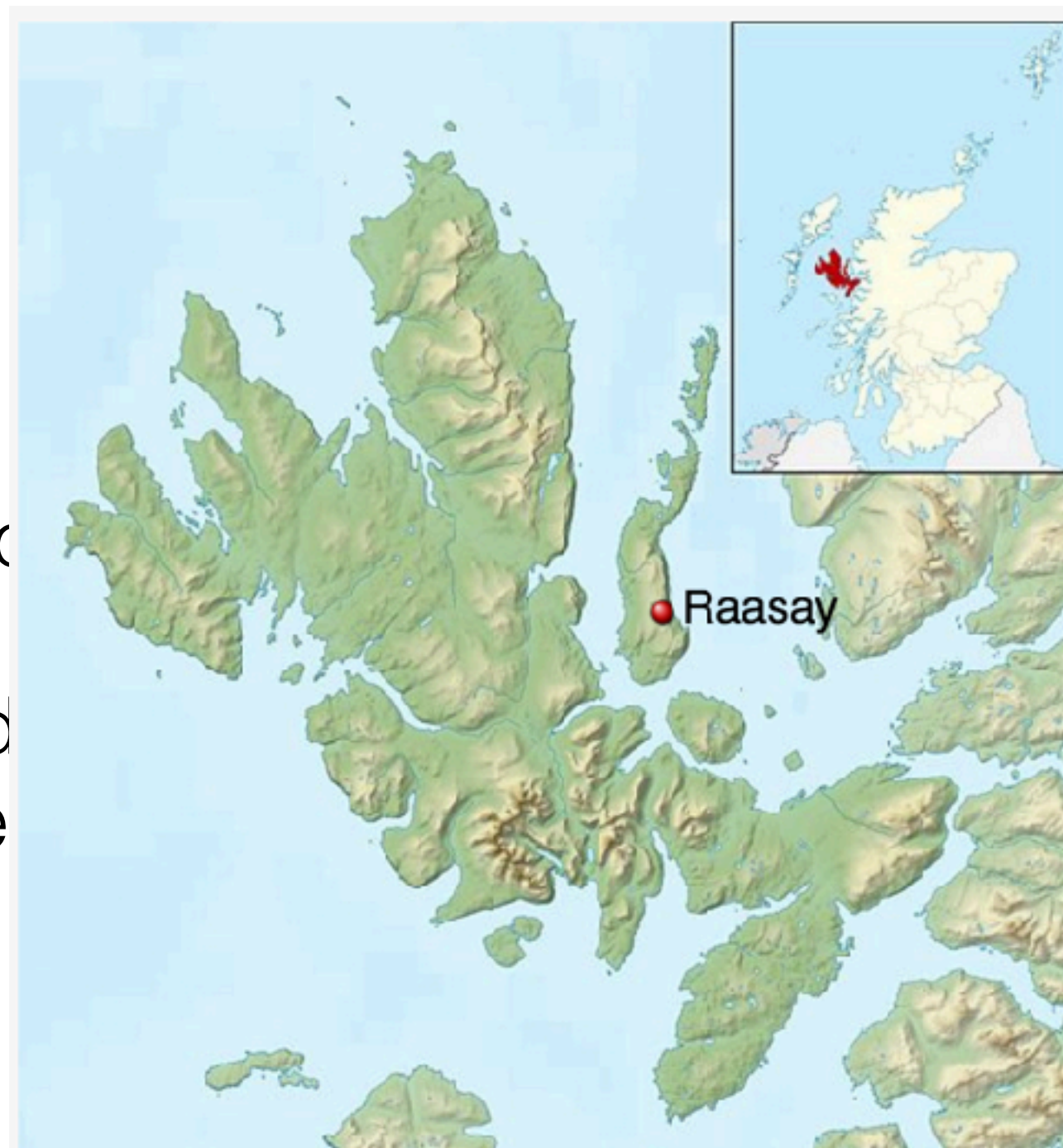
The Raasay Meeting & The Cambridge Workshop

Özgür Akgün
University of St Andrews

- The Raasay Meeting
 - Historical record linkage on the Isle of Skye: A colloquium for historians and computer scientists
 - 27 - 30 August 2016
- The Cambridge Workshop
 - Data Linkage: Techniques, Challenges and Applications
 - 12 - 16 September 2016

Raasay

- Small isle
- "The size of a hand of 161."
- Part of a o transcribe



The Mechanical Demographer

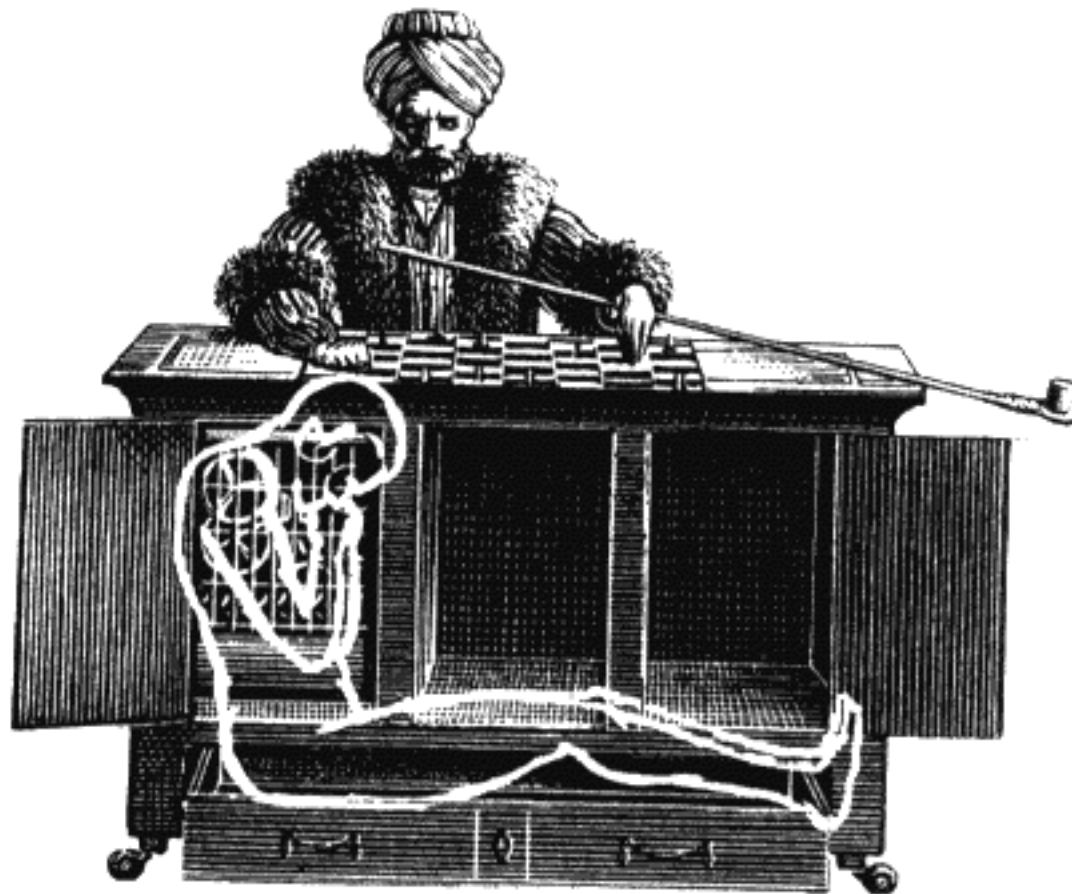


Plate 3.

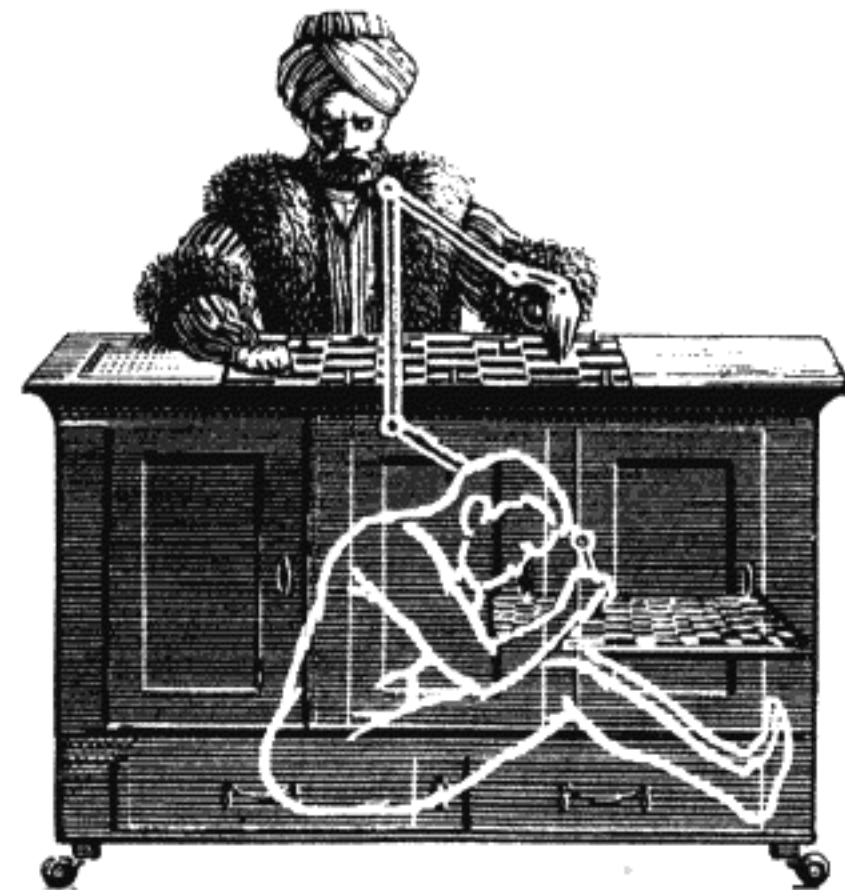


Plate 4.

Raasay - Participants

- Peter Christen
- Graham Kirby, Al Dearle, Tom Dalton, Özgür Akgün
- Chris Dibben
- Diego Ramiro
- Eilidh Garrett, Christine Jones, Lee Williamson
- Julia Jenning
- Alasdair Gray, Ahmad Al-Sadeeqi













Raasay - Angles

- Computer Science
 - Statistical Linkage
 - Rule-based Linkage
 - Synthesising Data & Error Injection
- Historical demography
 - How they do linkage "by hand"

Statistical Linkage

- Similarity metrics
- Calculate between pairs of records
- *Blocking* to avoid exhaustive calculation of all pairs
- Link if highly similar, non-link if highly dissimilar
- In the middle, clerical review

Statistical Linkage

- Downsides: a lot of things to tweak
 - Blocking
 - Fields to use
 - Similarity metric
 - Local decisions, only looking at a pair
- Upsides: computation speed

Statistical Linkage

- Advanced techniques
- Clustering based approaches
 - Instead of the simple threshold technique
- Multiple sources
- Privacy preserving

Rule-based Linkage

- If a condition holds, designate a link
- Tony Wrigley started compiling rules in 1973 in his book "Identifying people in the past"
- Automated discovery of rules: rule mining
- Upside: Easier to describe why a link is made

Synthesising Data

- Mainly for evaluating linkage techniques
- Real data lacks "ground truth" for linkage evaluation
- Statistically validated

Error Injection

- Synthesised data is error-free
- No missing data, typos, transcription errors
- Everybody wants a error-rate vs linkage-quality plot
- Even with real data

Linkage "by hand"

- Not really by hand, they use a similar process
- Normalisation, database queries
- They focus on family structures, starting from censuses

Using linked data

- Age specific marital fertility rates
 - In the 1880s young women in Skye were having a very large number of babies
 - >550 births per 1000 married woman per year
- Who are the front runners on certain behaviours?
- Richer people get married in England
- Use the ratio of household / number of servant as a way of partitioning the population

Orkney - Julia Jennings

- Very focused, very detailed
- 6 northernmost islands in Orkney
- Started ~15 years ago
- Digitised historical maps, archaeological surveys, etc
- Aerial photos: down to sheep tracks
- Finished linking BMD within parishes, and C to C.
 - Not C to BDM yet
- Economic indicators (price of barley) vs timing of death

Identifiers

- How to refer to individual records and people on records
- Consistently, across different research groups
- Alasdair Gray: Common scheme for identifiers

- Al Dearle on infrastructure
- Graham Kirby on identifying (encoding) multiple causes of death from a free form string
- Tom on synthesising data
- Ahmad on evaluating linkage algorithms
- Özgür on Combinatorial Optimisation

Upcoming

- November meeting in St Andrews
 - Working Together: CS + Historians
- Explorathon (European Researchers' Night)
 - September 30th in St Andrews
- SICSA DEMOfest
 - November 11th in Glasgow

Cambridge Workshop

- Administrative Data Linkage
 - Germany, Netherlands, US
- Researchers (CS/Statistics/Applications)
- Tool developers
- Data suppliers
- <http://www.newton.ac.uk/event/dlaw02>

