

Reconstructing Genealogical Relationships from Public Registers

Graham Kirby, Alan Dearle, Masih Hajiarab Derkani, Tom Dalton, Konstantinos Daras
University of St Andrews

Chris Dibben, Lee Williamson, Zengyi Huang, Zhiqiang Feng
University of Edinburgh

Eilidh Garrett, Alice Reid
University of Cambridge



Background

- Digitising Scotland project
 - transcribe vital events records 1855-1973
 - births
 - deaths
 - marriages
 - codify occupations, causes of death, locations
 - link records to form family tree(s)
- [*www.lscs.ac.uk/projects/digitising-scotland/*](http://www.lscs.ac.uk/projects/digitising-scotland/)

Why?

- To produce resource for further research
 - patterns of migration within Scotland
 - social mobility
 - impact of developing welfare state
 - impact of industrialisation
 - distributions of occupations
 - effects of occupation on health
 - long-term impacts of serious viruses
 - genetic influences on disease
 - ...

Scottish Public Records

- Recorded nationally since 1855
- Birth records: 14 million
- Marriage records: 4 million
- Death records: 11 million
- About 18 million individuals

1868. BIRTHS in the Central of District in the Borough of Glasgow

(1.) (2.) (3.) (4.) (5.) (6.)

No.	Name and Surname.	When and Where Born.	Sex.	Name, Surname, & Rank or Profession of Father. Name, and Maiden Surname of Mother. Date and Place of Marriage.	Signature and Qualification of Informant, and Residence, if out of the House in which the Birth occurred.	When and Where Registered, and Signature of Registrar.
1360	Charles Rennie McIntosh	1868 June Percents 11th 15m A.M.	M.	William McIntosh Police Clerk Margaret McIntosh M. S. Rennie 71 Parson Street Glasgow	W. McIntosh Father Present	1868 June 25th at Glasgow Arch. Hood Asst. Registrar SD
1361	Rae Ann Caldwell	1868 May Thirtieth 1st 15m A.M.	F.	William Caldwell Iron Moulder Mary Caldwell M. S. McChaffrey 2 North Street Glasgow	Mary Caldwell Mother	1868 June 25th at Glasgow Arch. Hood Asst. Registrar SD
1362	Walter Dalgliesh Jackson	1868 June Eighteenth 1st 10m P.M.	M.	Walter Dalgliesh Jackson Railway Clerk Jane Jackson M. S. Dunlop 68 North St. Mungie Glasgow	M. S. Jackson Father Present	1868 June 26th at Glasgow Arch. Hood Asst. Registrar SD

B

James Davidson Registrar

BIRTHS in the Central of District in the 1868

(1.) (2.) (3.) (4.)

No.	Name and Surname.	When and Where Born.	Sex.	Name, Surname, & Rank or Profession of Father. Name, and Maiden Surname of Mother. Date and Place of Marriage.	S.
1360	Charles Rennie	1868	M.	William McIntosh	
	McIntosh	June		Police Clerk	
		Perth			
		11 h 15 m A.M.		Margaret McIntosh	
				W. L. Rennie	
		70 Parson Street			
		Glasgow		1862 August 4th Glasgow	
	Rose Ann	1868	F.	William Caldwell	
	Caldwell	May		Iron Moulder	
		21 1 10			

(Page 62.)

1880. MARRIAGES in the District of Hamilton in the County of Fawk

No.	When, Where, and How Married.	Signatures of Parties. Rank or Profession whether Single or Widowed, and Relationship (if any).	Age.	Usual Residence.	Name, Surname, and Rank or Profession of Father. Name, and Maiden Surname of Mother.	If a regular Marriage, Signatures of officiating Minister and Witnesses. If irregular, Date of Concoction, Decree of Declarator, or Sheriff's Warrant.	When & Where Registered, and Signature of Registrar.
123	1880. on the <u>Third</u> day of <u>August</u> at <u>8 of Rowmers</u> <u>Hamilton</u>	<u>Signed,</u> <u>Jamies Mack Hardie</u> <u>Miners Secretary</u> <u>Bachelors</u>	<u>23</u>	<u>Rowmers</u> <u>Buildings</u> <u>Rowmers</u> <u>Hamilton</u>	<u>William Aithen</u> <u>Coal Miner</u> <u>Mary Hardie</u> <u>M. S. Kerr</u>	<u>Signed,</u> <u>George Bell</u> <u>E. W. Minister</u> <u>Ester B. Mcmason</u> <u>W. R. Minister</u> <u>(Signed)</u>	1880, <u>August 6th</u> <u>Hamilton</u>
	<u>After Banns</u> <u>according to the forms</u> <u>of the United Presbyterian</u> <u>Church and Evangelical</u> <u>Union, Church</u>	<u>Signed,</u> <u>Lilias Wilson</u> <u>Dressmaker</u> <u>(Spinster)</u>	<u>17</u>	<u>Rowmers</u> <u>Hamilton</u>	<u>Duncan Wilson</u> <u>Coal Miner</u> <u>Sarah Wilson</u> <u>M. S. Miller</u>	<u>William Wilson</u> <u>Witness</u> <u>Andrew Forrest</u> <u>Russell</u> <u>Witness</u>	<u>William Wilson</u> <u>Asst. Registrar.</u> <u>M</u>
124	1880. on the <u>Fifth</u> day of <u>August</u> at <u>First House</u> <u>Hamilton</u>	<u>Signed,</u> <u>John Bladen</u> <u>his & month</u> <u>A. W. Hamilton</u> <u>Coal Miner</u> <u>Bachelors</u>	<u>34</u>	<u>25 Whitehill</u> <u>Road</u> <u>Bumboth</u> <u>Hamilton</u>	<u>James Bladen</u> <u>Coal Miner</u> <u>Mary Bladen</u> <u>M. S. Wood</u> <u>(Both deceased)</u>	<u>Signed,</u> <u>Henry W. Hamilton</u> <u>Minister at Hamilton</u>	1880, <u>August 7th</u> <u>Hamilton</u>
	<u>After Publication</u> <u>according to the forms</u> <u>of the Church of</u> <u>Scotland</u>	<u>Signed,</u> <u>Margaret Watson</u> <u>her & month</u> <u>A. W. Hamilton</u> <u>(Widow)</u>	<u>34</u>	<u>25 Whitehill</u> <u>Road</u> <u>Bumboth</u> <u>Hamilton</u>	<u>James Docherty</u> <u>Coal Miner</u> <u>Mary Docherty</u> <u>M. S. Robertson</u> <u>(Both deceased)</u>	<u>Signed,</u> <u>James H. Ingham</u> <u>Witness</u> <u>Margaret M. Ingham</u> <u>her month & witness</u> <u>A. W. Hamilton</u> <u>Witness</u>	<u>William Hoyer</u> <u>Asst. Registrar.</u> <u>M</u>

William Hoyer Registrar.

1880. DEATHS in the District of St. Mary in the Burgh of Quindoe Page 47-

No.	Name and Surname. Rank or Profession, and whether Single, Married, or Widowed.	When and Where Died.	Sex.	Age.	Name, Surname, & Rank or Profession of Father. Name, and Maiden Surname of Mother.	Cause of Death, Duration of Disease, and Medical Attendant by whom certified.	Signature & Qualification of Informant, and Residence, if out of the House in which the Death occurred.	When and where Registered, and Signature of Registrar.
139	George McIntosh. Railway Guard. Married to Isabella Butters.	1880, February Sixteenth. Found at Ethieharum Inverkeithing. About 8h. A.M. No. 11-25 Hawthick Quindoe.	Male	43 Years	George Mcintosh, Master. Martha Mcintosh, No. 1, McCarlane (Deceased).	Accidentally Drowned from fall of a Railway Train and portion of Tay Bridge into the North Sea, on December 28, 1879. Not certified.	Geo. McIntosh Father, 121 Hawthick.	1880, February 17 th At Quindoe. Jas. Anderson, Registrar.
140	James McIntosh. Watchmaker (Jewelleryman). (Illegitimate). Married to Elizabeth McMaster.	1880, February Sixteenth. 8h. 20m. P.M. 179 Lochee Road Quindoe.	Male	20 Years	Robert Mcintosh, Household Labourer. Ann Luthersland, Dunelmee Weaver (Deceased).	Consumption. 6 Months. Infect. by David Macdonald Surgeon.	Robert Mcintosh Father, 1 South Mid Street Lochee. (Present).	1880, February 17 th At Quindoe. Jas. Anderson, Registrar.
141	David Jobson. Pub. Merchant. Married to Mary Watson Miller.	1880, February Seventeenth. Found in River Tay, about 10h. A.M. No. 11-3 Millie Place, Quindoe.	Male	41 Years	David Jobson, Hire Merchant (Deceased). Mary Jobson, No. 1, Quindoe (Deceased).	Accidentally Drowned from fall of a Railway Train and portion of Tay Bridge into the North Sea, on 28, 1879. Not certified.	James Jobson Brother, 10, Harrington Terrace, Broughty Ferry.	1880, February 17 th At Quindoe. Jas. Anderson, Registrar.

D

Jas. Anderson, Registrar.

Problem: Codification

Problem: Codification

- Classify causes of death to ICD10 (about 1500 codes)

Problem: Codification

- Classify causes of death to ICD10 (about 1500 codes)
 - “cerebral haemorrhage”

Problem: Codification

- Classify causes of death to ICD10 (about 1500 codes)
 - “cerebral haemorrhage”
 - “cancer of colon”

Problem: Codification

- Classify causes of death to ICD10 (about 1500 codes)
 - “cerebral haemorrhage”
 - “cancer of colon”
 - “heart failure”

Problem: Codification

- Classify causes of death to ICD10 (about 1500 codes)
 - “cerebral haemorrhage”
 - “cancer of colon”
 - “heart failure”
 - “effusion on the brain caused by excessive drinking and exposure to cold”

Problem: Codification

- Classify causes of death to ICD10 (about 1500 codes)
 - “cerebral haemorrhage”
 - “cancer of colon”
 - “heart failure”
 - “effusion on the brain caused by excessive drinking and exposure to cold”
 - “drowned by a smack on boat having been culpably negligently or recklessly navigated or steered by Angus McPherson boatman, Portree”

Problem: Codification

- Classify causes of death to ICD10 (about 1500 codes)
 - “cerebral haemorrhage”
 - “cancer of colon”
 - “heart failure”
 - “effusion on the brain caused by excessive drinking and exposure to cold”
 - “drowned by a smack on boat having been culpably negligently or recklessly navigated or steered by Angus McPherson boatman, Portree”
 - “drowning possibly through his falling asleep or getting insensible through liquor on the sea shore at low water and being afterwards covered by the sea on the tide flowing”

Problem: Tree Reconstruction

- From the set of certificates, need to work out:
 - how many people?
 - what were the relationships?

Problem: Tree Reconstruction

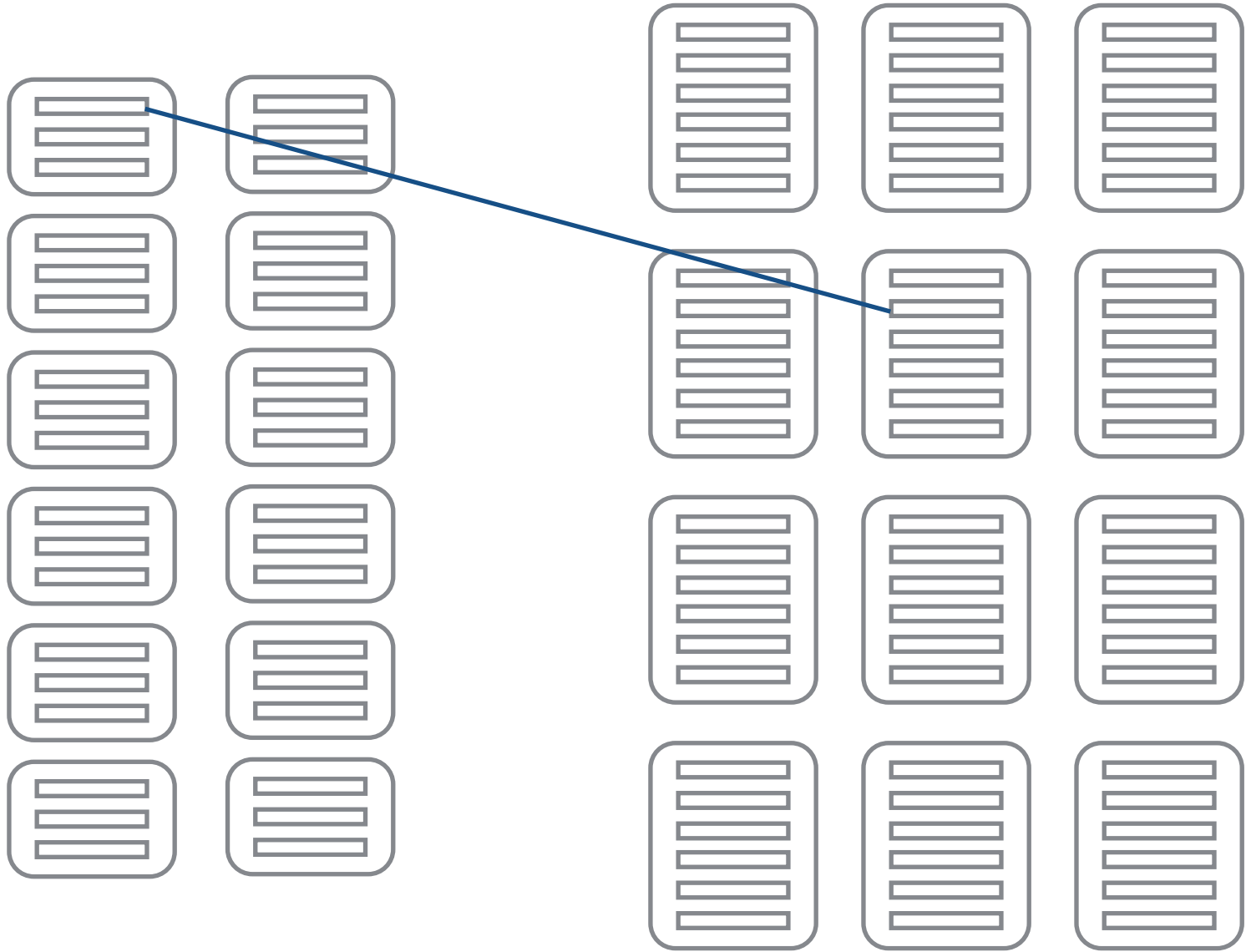
- Match up information on certificates
 - may be ambiguous
 - information may be missing
 - mistakes
 - immigration and emigration

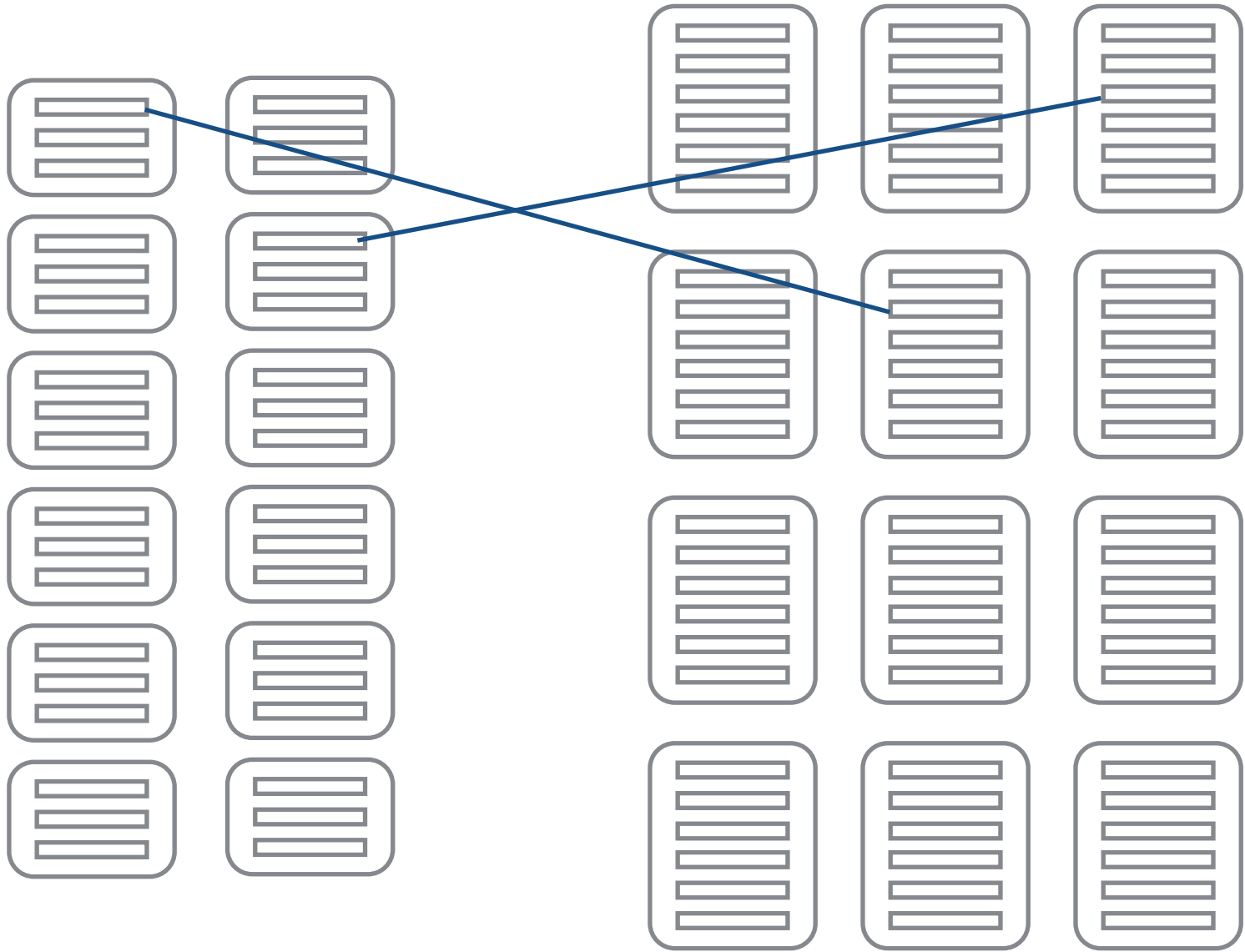
Problem: Tree Reconstruction

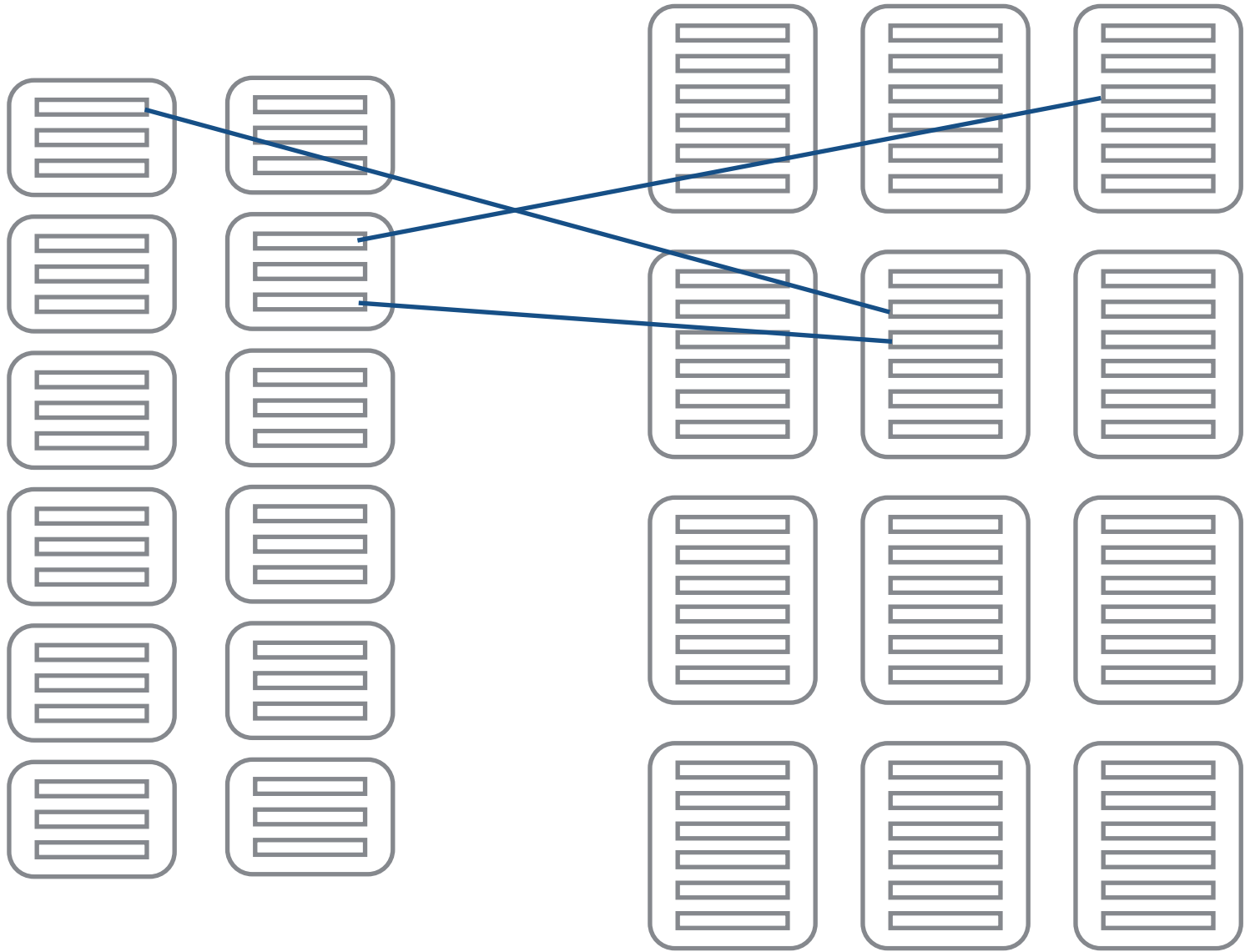
- Birth Certificate 3826756
 - Jane Smith, born 1/3/1889
- Marriage Certificate 573734
 - Janet Smyth, born 11/3/1889
- How many people?
- Also consider other information e.g. place of birth, occupation, relationships to other people

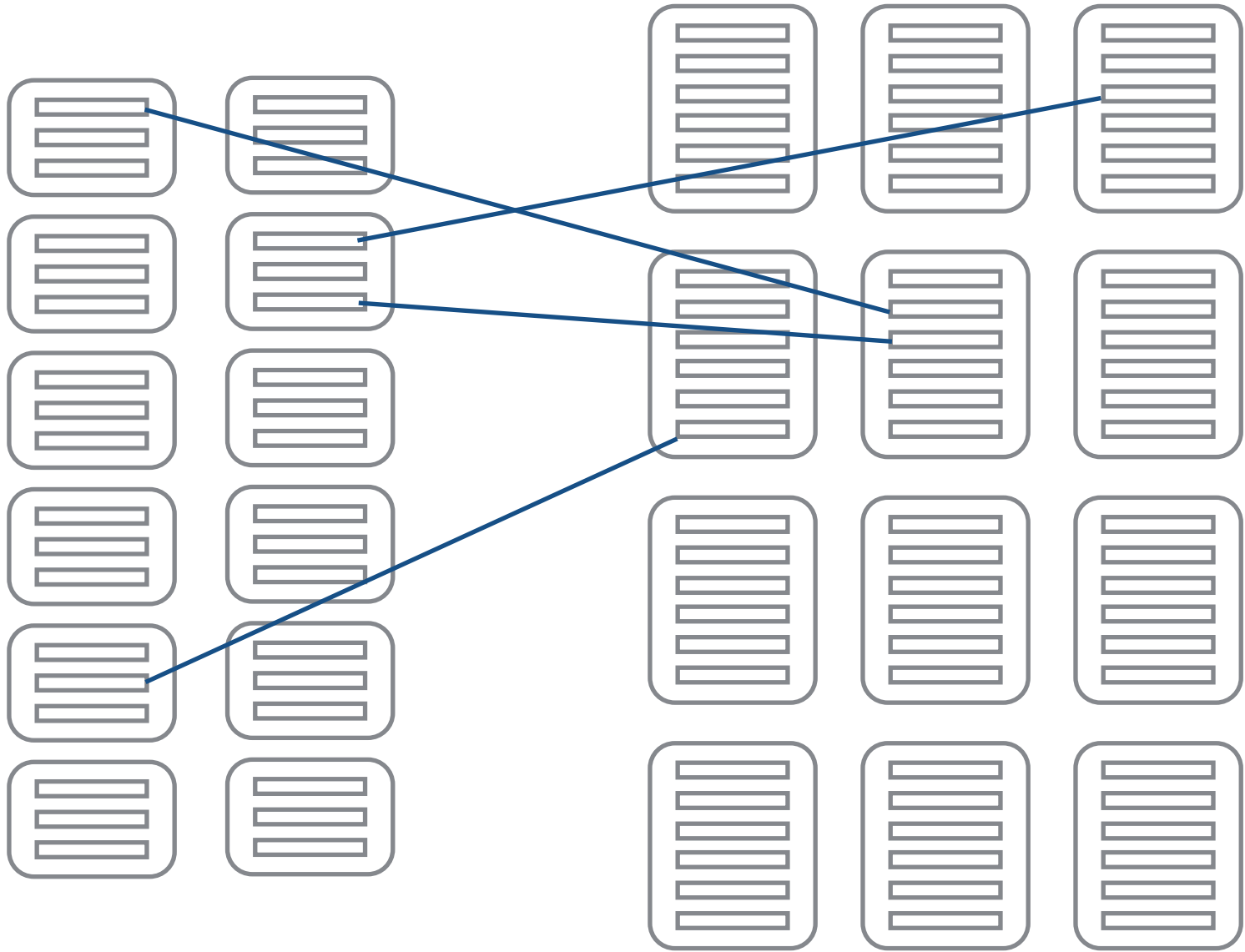


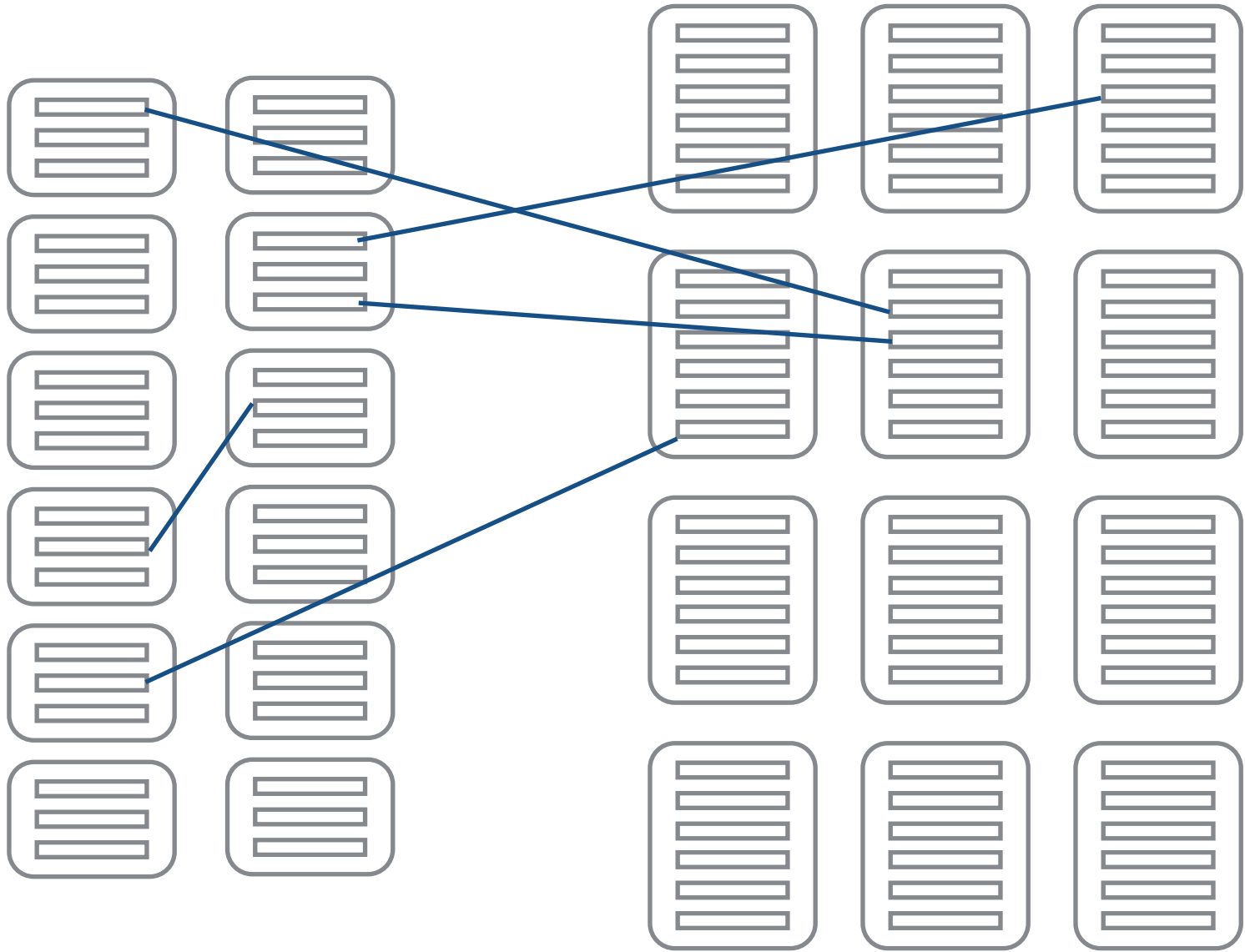


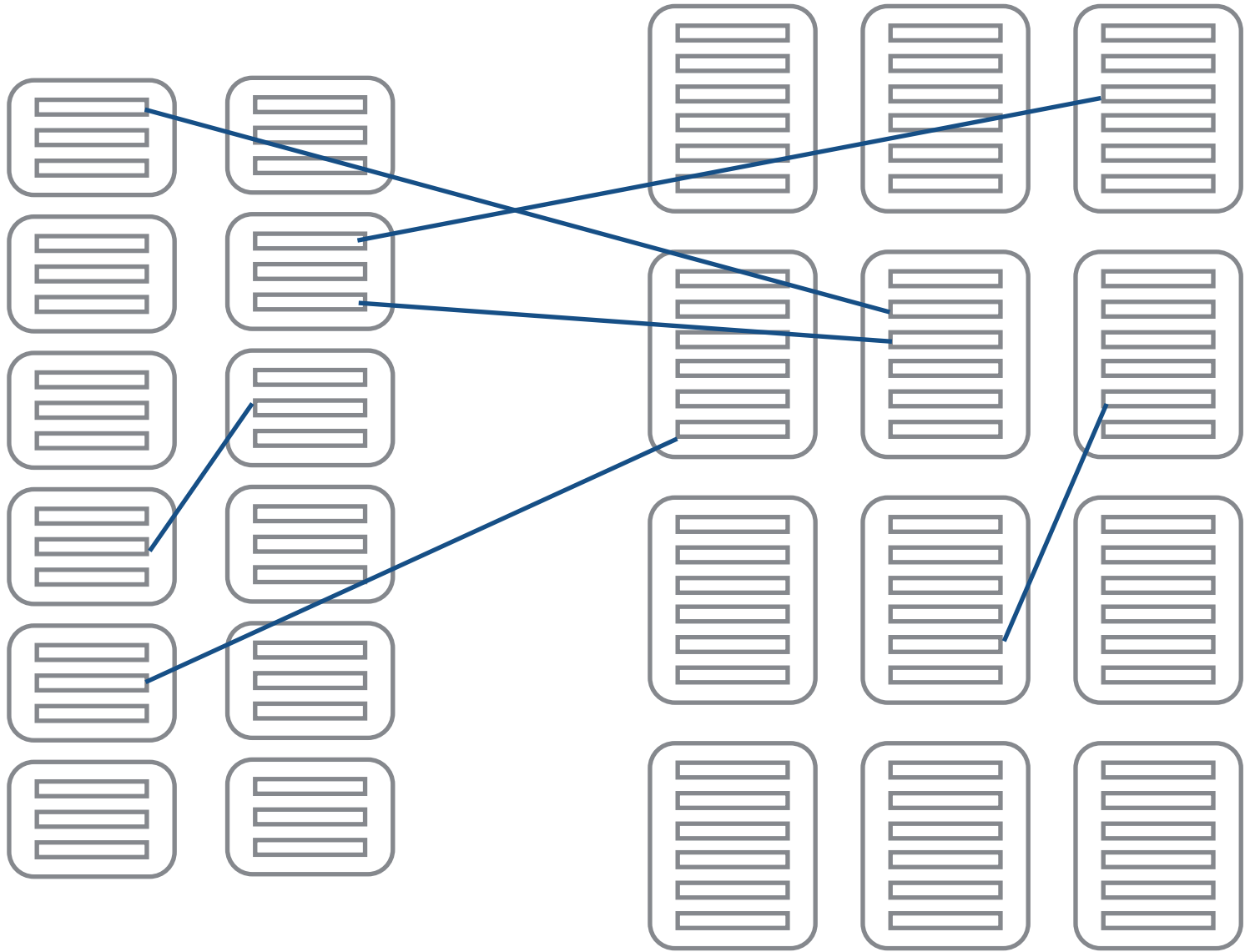












Problem: Scale

Problem: Scale

- Even assuming you can decide whether a given pair matches, there are a lot of possible pairs to match...

Problem: Scale

- Even assuming you can decide whether a given pair matches, there are a lot of possible pairs to match...
- 3 people mentioned on a birth certificate

Problem: Scale

- Even assuming you can decide whether a given pair matches, there are a lot of possible pairs to match...
 - 3 people mentioned on a birth certificate
 - 6 on a marriage certificate

Problem: Scale

- Even assuming you can decide whether a given pair matches, there are a lot of possible pairs to match...
 - 3 people mentioned on a birth certificate
 - 6 on a marriage certificate
- $3 \times 14 \text{ million} \times 6 \times 4 \text{ million} = 10^{15}$

Problem: Scale

- Even assuming you can decide whether a given pair matches, there are a lot of possible pairs to match...
 - 3 people mentioned on a birth certificate
 - 6 on a marriage certificate
- $3 \times 14 \text{ million} \times 6 \times 4 \text{ million} = 10^{15}$
- If we could check a pair in a microsecond this would take 32 years

Problem: Scale

- Even assuming you can decide whether a given pair matches, there are a lot of possible pairs to match...
 - 3 people mentioned on a birth certificate
 - 6 on a marriage certificate
- $3 \times 14 \text{ million} \times 6 \times 4 \text{ million} = 10^{15}$
- If we could check a pair in a microsecond this would take 32 years
- Also need to check births against births & deaths, deaths against deaths, births, marriages, ...

Problem: Evaluation

Problem: Evaluation

- How do you know whether the reconstructed tree is correct?

Problem: Evaluation

- How do you know whether the reconstructed tree is correct?
- In general, you can't: very little ground truth
 - the whole point of doing this automatically is that we don't have time for experts to do it
 - but we can get experts to look at some parts

Problem: Evaluation

- How do you know whether the reconstructed tree is correct?
- In general, you can't: very little ground truth
 - the whole point of doing this automatically is that we don't have time for experts to do it
 - but we can get experts to look at some parts
- For intended research purposes, doesn't matter as long as it's reasonably close

Problem: Evaluation

- How do you know whether the reconstructed tree is correct?
- In general, you can't: very little ground truth
 - the whole point of doing this automatically is that we don't have time for experts to do it
 - but we can get experts to look at some parts
- For intended research purposes, doesn't matter as long as it's reasonably close
- Also possible to evaluate linkage techniques with synthetic data

Problem: Understanding Results

Problem:

Understanding Results

- Sometimes we won't be able to make good automatic matching decisions
- need to record this uncertainty and let researchers know about it

Problem:

Understanding Results

- Sometimes we won't be able to make good automatic matching decisions
 - need to record this uncertainty and let researchers know about it
- How to present this in a way that's easy to understand?
 - multiple possible family trees for a given person

Summary

- **Aim**
 - produce linked data set containing all people born, married or died in Scotland 1855-1973
- **Problems**
 - transcription of data
 - classification of occupations and causes of death
- **Research issues**
 - techniques for linking accurately and efficiently
 - understanding the consequences of uncertainty