

Reconstructing Genealogical Relationships from Public Registers

Graham Kirby, Alan Dearle
University of St Andrews

Chris Dibben, Lee Williamson
University of Edinburgh



University of
St Andrews



THE UNIVERSITY
of EDINBURGH

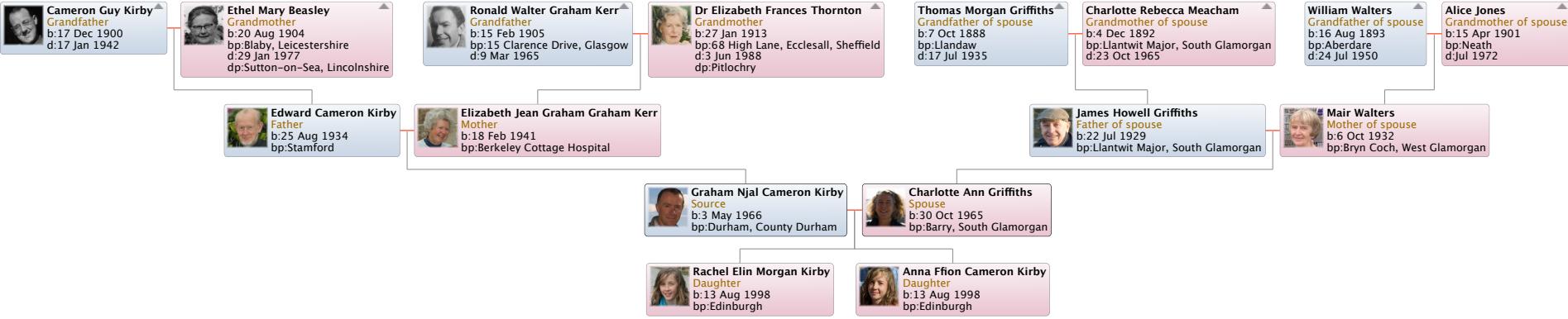


Background

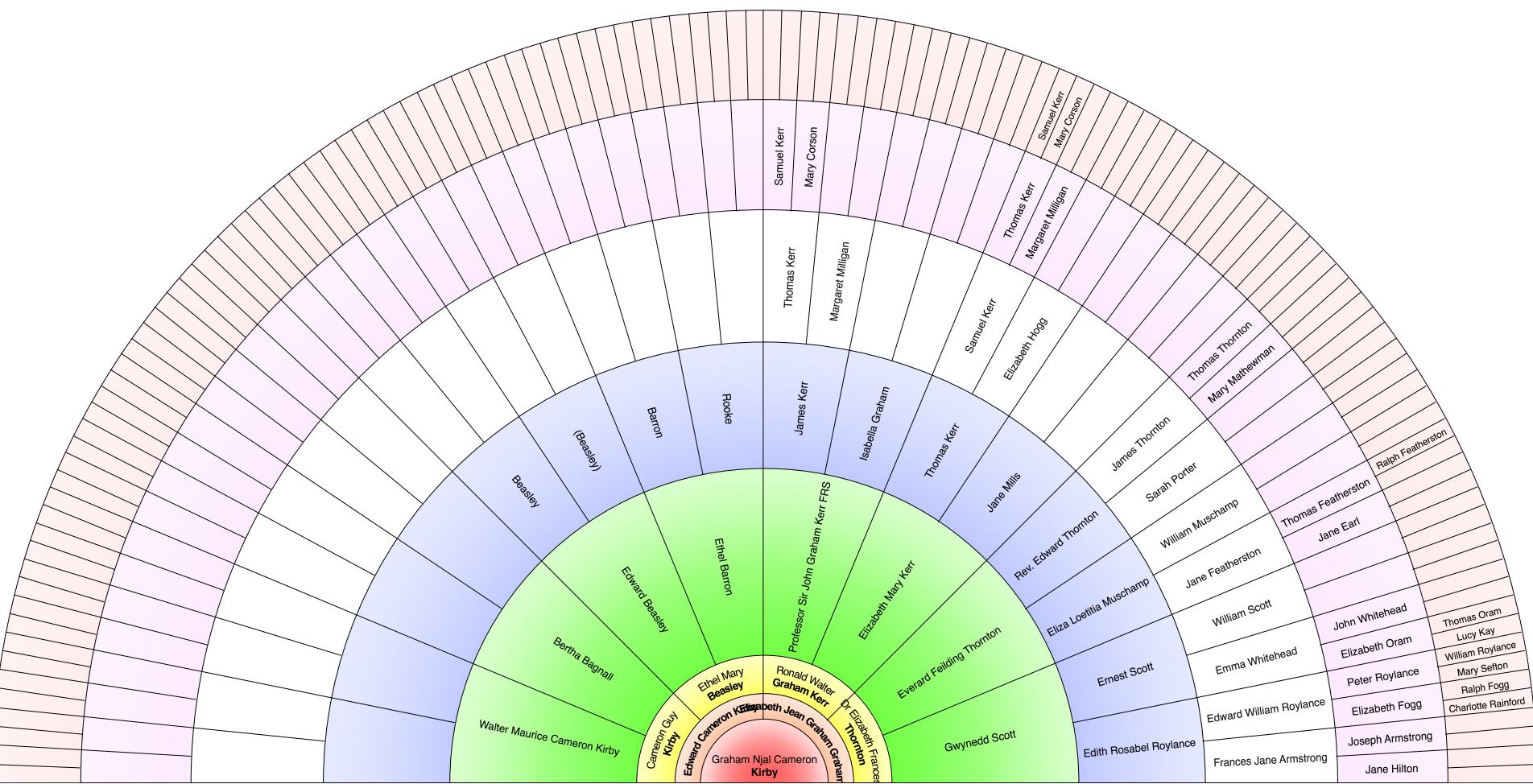
- Digitising Scotland project
 - Transcribe vital events records 1855-1973
 - births
 - deaths
 - marriages
 - Link records to form family tree(s)

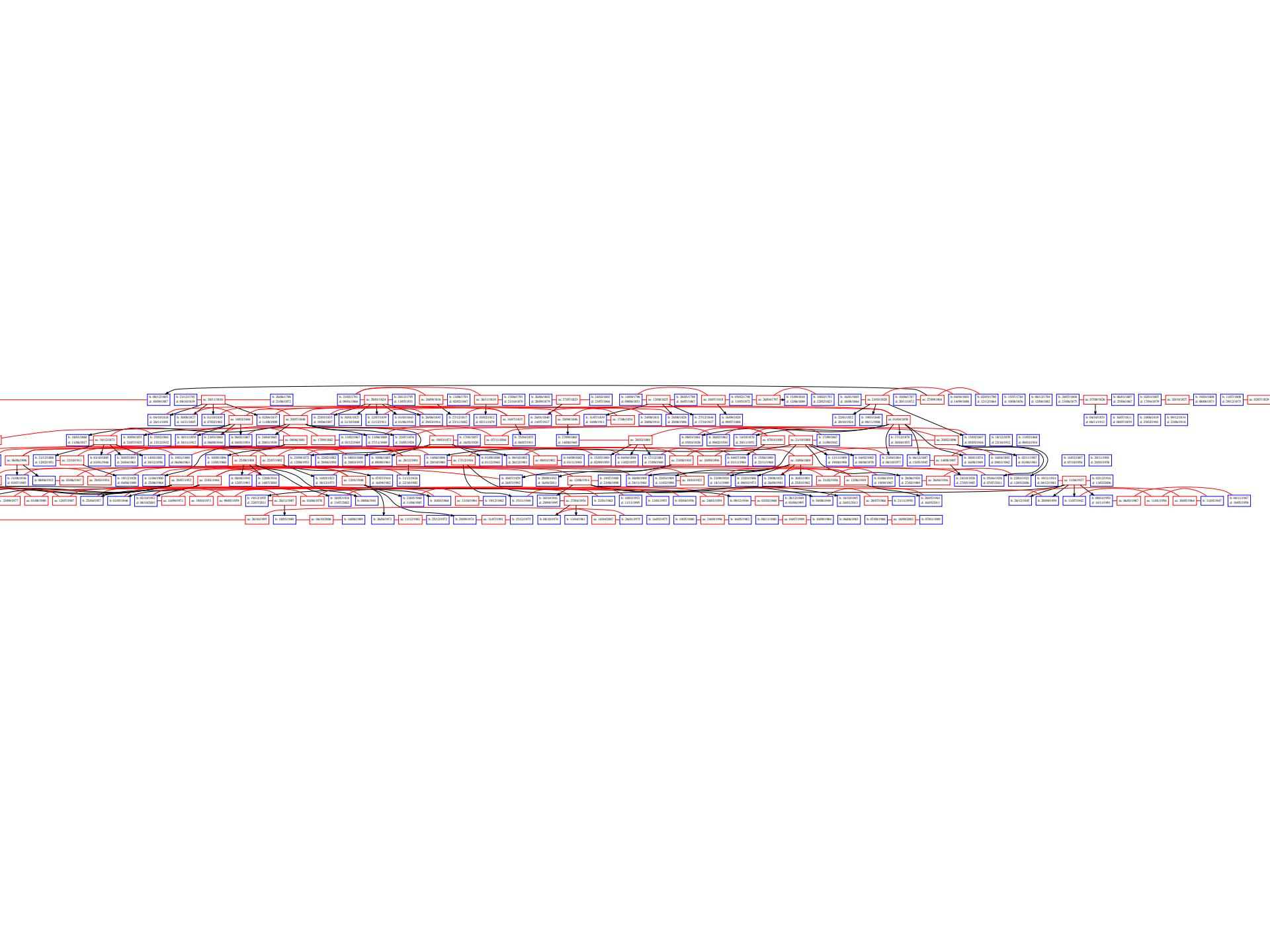
Why?

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









Scottish Public Records

	Records	Estimated Size
Births	14 million	1.9 GB
Deaths	11 million	2.3 GB
Marriages	4.2 million	0.8 GB

- About 18 million individuals

Problem: Data Acquisition

- Data exists as scanned images
- Need to extract the text so it can be processed
- Transcription: image to text

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

(1.)	(2.)	(3.)	(4.)	(5.)	(6.)
	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.
1860	Charles Annie McIntosh	1868 June Seventeenth 11h 15m A.M.	M.	Mr. William McIntosh Police Clerk Metaphab McIntosh W. L. Kenzie	W. McIntosh Father Present
	70 Parson Street Glasgow	1868 August 1 st	F.	Glasgow	1868 June 25 th at Glasgow Arch Hood Mid. Registrar
1861	Frederick Caldwell	1868 May Thirty-first 10 h 15 m A.M.	M.	William Caldwell Cross Moulder Mary Caldwell Mr. S. McSheffrey	Frederick Caldwell Mother 1868 June 25 th at Glasgow Arch Hood Mid. Registrar
1862	Walter Dalgliesh Jackson	1868 June Eighteenth 10 am P.M.	M.	Walter Dalgliesh Jackson Railway Clerk Jane Jackson Mr. S. Dunlop	W. Jackson Father Present 1868 June 26 th at Glasgow Arch Hood Mid. Registrar
	68 North St. Mungo That Glasgow	1868 June 15 th	F.	Glasgow	1868 June 15 th

Thomas Davidson Registrar

BIRTHS in the Central of District in the Town

(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.
	Charles Rennie McIntosh	1868 June Penzance 11 h 15 m A.M.	M	William McIntosh Police Clerk Margaret McIntosh W. J. Rennie
1360		70 Parson Street Glasgow		1862 August 4 th Glasgow
	Rose Ann Caldwell	1868 May 11 h 10 m	F	William Caldwell Iron Moulder

Birth Record - Transcribed

- From the transcription process we get a record like this:
- 356|JENNER|MARY|F|||||1866|
COULTHARD|||19|4||JAMES|0|WRITER IN
NAVY DOCKYARD|MARGARET|0||25|1|||
1843|||||||
- Birth record 356 for Mary Jenner...

1880. MARRIAGES in the District of Hamilton

In the County of Lanark

No.	When, Where, and How Mated.	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.	If a regular Marriage, Signatures of officiating Minister and Witnesses. If Irregular, Date of Conclusion, Name of Cleric, or Sheriff's Warrant.	When & Where Registered, and Signature of Registrar.
123	1880. on the Third day of August at 20 Rowntrees Hamilton	Signed, James Keir Hardie Business-Secretary Bracheton	23	Rowntree's Buildings Rowntrees Hamilton	William Aitken Coal Miner S. U. Minister Mary Hardie M. S. Kerr	George Bell S. U. Minister C. G. Morrison U. H. Minister (Signed)	1880. August 1st
	After Banns according to the Form of the United Presbyterian Church and Evangelical Union, Church	Signed, Lilias Wilson Dressmaker (Spinster)	17	Rowntrees Hamilton	Duncan Wilson Coal Miner Sarah Wilson M. S. Miller	William Miller Witness Andrew Forrest Russell Wilson	William Miller Post Register Miller
124	1880. on the Fifth day of August at First House Hamilton	Signed, John Gladden his x mark 8, Mr. Hamilton Street Coal Miner Bracheton	34	25 Whitehill Road Bumbank Hamilton	James Gladden Mary Gladden J. S. Wood (both deceased)	Signed, Henry M. Hamilton Minister at Hamilton Hamilton	1880. August 9th
	After Publication according to the Form of the Church of Scotland	Signed, Margaret Watson her x mark 8, Mr. Hamilton Street (widow)	34	25 Whitehill Road Bumbank Hamilton	James Docherty Coal Miner Mary Docherty J. P. Robertson (both deceased)	James Watson Margaret M. Anstruther for mother Watson H. M. Hamilton Witness	William New Post Register Mr.

William Scott

Registrar.

1880. MARRIAGES in the District of Stornoway

No.	When, Where, and How Married.	Signatures of Parties. Rank or Profession whether Single or Widowed, and Relationship (if any).	Age.
123	<p>1880, on the <u>Third</u> day of <u>August</u> at <u>85 Cowmases</u> <u>Harrowton</u></p>	<p>Signed, James Keir Hardie Minister-Secretary (Bachelor)</p>	<p>66 23 20 66</p>
	<p>After Banns according to the Form of the United Presbyterian Church and Evangelical Methodist Church</p>	<p>Signed, Lilias Wilson Dressmaker (Spinster)</p>	<p>86 17 3</p>

1880. DEATHS in the District of Aberdeen in the Burgh of Aberdeen

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, Surname, & Rank or Profession of Mother.	Cause of Death, Duration of Disease, and Medical Attendant by whom certified.	Signature & Qualifications of Informer, and Residence, if out of the House in which the Death occurred.	When and where Registered, and Signature of Registrar.
130	George Mr. McIntosh Railway Guard.	1880. February Sixteenth. Found at Isobell's Barn Inverkeillan.	Male	43	George McIntosh, Carters.		Accidentally drowned falling off a railway train and perishing in Tay Bridge into the River Tay, av.	Geo. McIntosh Brother 181 Knockhill, At Dundee.	1880. February 17 th At Dundee.
130	Married to Isabella Buttress.	about 2h. A.M. Aberdeen.			Mr. McIntosh	Ms. L. McFarlane	Decidedly ill. December 22, 1879.		Jas. Anderson Registrar.
140	James Mr. McIntosh Matchmaker Gourymann (Illegitimate)	1880. February Sixteenth. 8h. 20m. P.M. 179 Inchicore Road	Male	20	Robert McIntosh Foundry Labourer.	Ann Letherland, Caucaloon Weaver	Consumption. 6 months. Is beat by David Alexander (Decreasd.)	Robert McIntosh Brother 1 South Mid Street, At Dundee.	1880. February 17 th At Dundee.
140	Advised to Elizabeth Mc Martin	Decreasd.					Surgeon.	Jas. Anderson Registrar.	
141	David Mr. Jobson Oil Merchant.	1880. February Sixteenth. about 10h. A.M. Found in Groatay	Male	21	David Johnson Stair Merchant	Mary Johnson Ms. Lender	Accidentally drowned falling off a railway train and perishing in Tay Bridge into the River Tay, Decreasd.	James Johnson Brother 10, Union Street, At Dundee.	1880. February 17 th At Dundee.
	Married to Mary Paterson Miller	Mr. & Mrs. Lender Place, Dundee.					Decreasd.		Jas. Anderson Registrar.

	MARRIED TO Elizabeth McMaster	DUNDEE.	
	David	1880.	Mr. H
	Jobson.	February Seventeenth.	Year
141	Bil Merchant.	Found in River Tay, about 10 h. M.	
	MARRIED TO Mary Paterson Miller	Mr. & Mrs. - 3 Shirle Place, Dundee.	

Problem: Classification

- 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

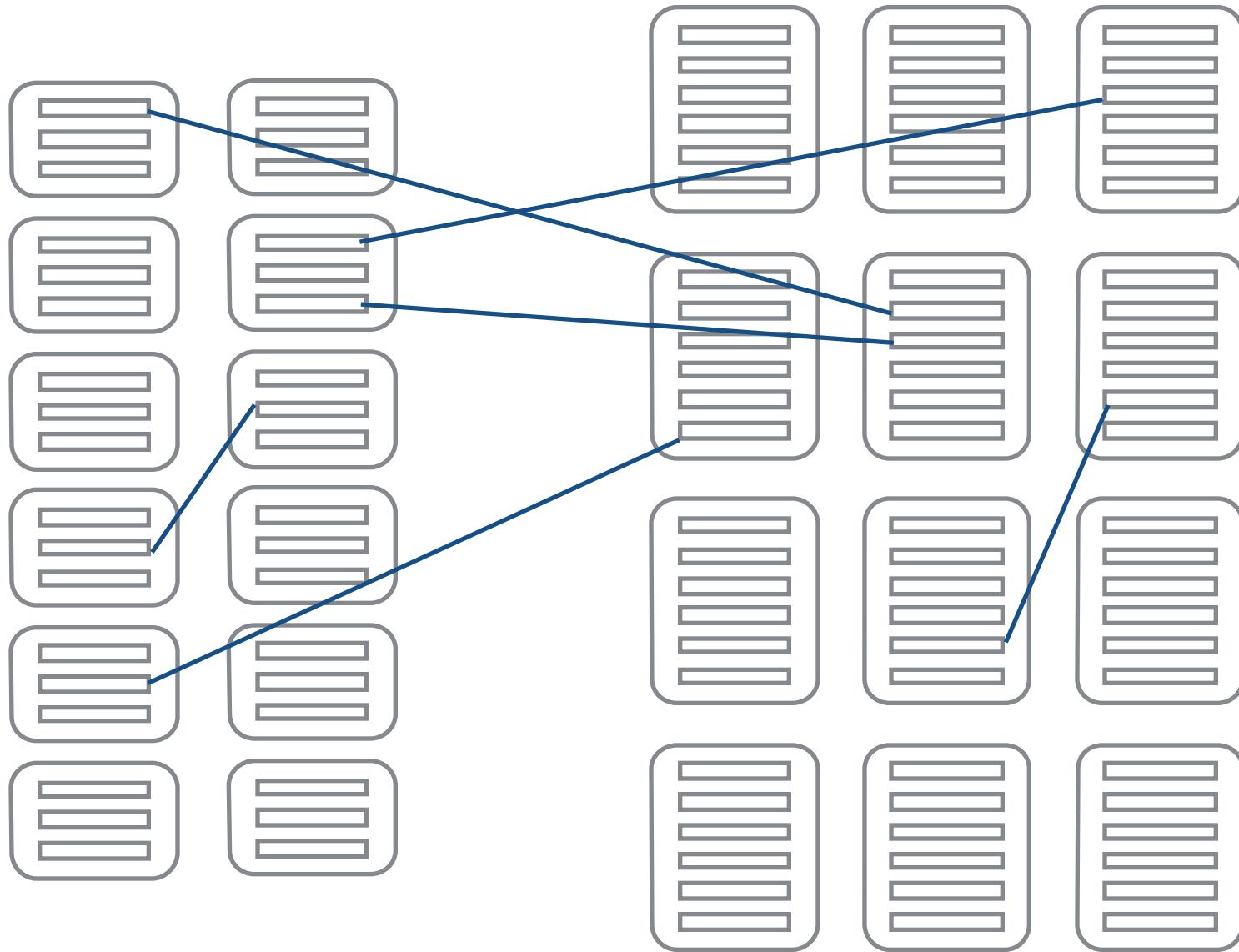
- From the set of certificates, need to work out:
 - how many people there were
 - what were the relationships between them

Problem: Tree Reconstruction

- Birth Certificate 3826756
 - Jane Smith, born 1/3/1889
- Marriage Certificate 573734
 - Janet Smyth, born 11/3/1889
- How many people?

Problem: Tree Reconstruction

- Match up information on certificates
 - may be ambiguous
 - bits may be missing
 - mistakes
 - people moving to or away from Scotland



Problem: Scale

- Even assuming you can decide, 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: Evaluating Correctness

- How do you know whether the reconstructed tree is correct?
- In general, you can't
 - 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

- 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