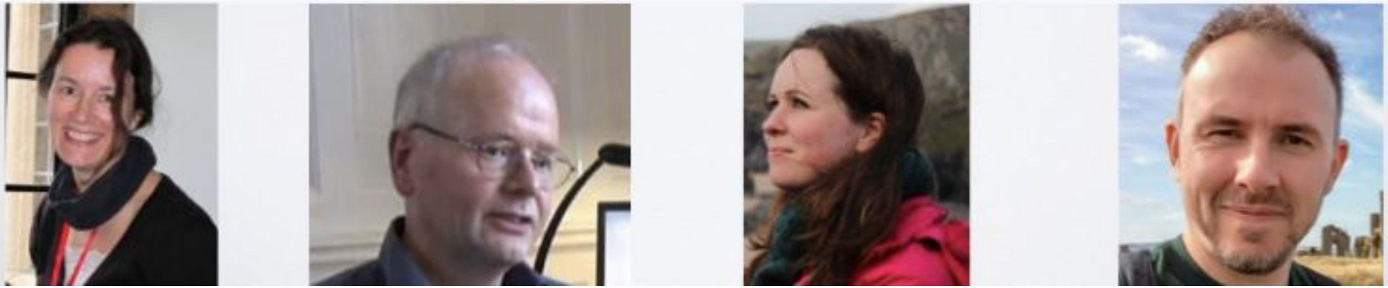
Mining the Stuart State Papers for Places

Yann Ryan,
University of Aberdeen, 10 November 2021.

Top: Sebastian Ahnert, Ruth Ahnert, Arno Bosse, Howard Hotson

Bottom: Miranda Lewis, Philip Beeley, Esther van Raamsdonk, and Matthew Wilcoxson



The Networking Archives Project

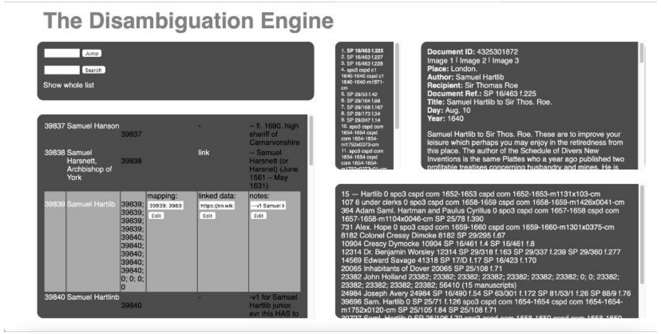
The project (PI Howard Hotson) provides a working model for a meta-archive, by bringing together:

- Existing content from Early Modern Letters Online (EMLO)
- The letters from the Tudor State Papers (supplied by State Papers Online, SPO), already cleaned as part of AHRC-funded project Tudor Networks of Power (PI R. Ahnert)
- A further body of letters from the Stuart State Papers (SPO II and III, 1603-1714), now in the final stages of being cleaned
- Totalling approx. 450,000 letters

Ultimately, this meta-archive is imagined as a central repository for early modern correspondence metadata more generally, which can be added to and analysed using a suite of tools that we have developed for data cleaning, reconciliation, and analysis.

Networking Archives and the State Papers

- Networking Archives team has undertaken extensive data cleaning of the State Papers Online, Stuart (1603-1714)
- People reconciliation (removing duplicates and splitting out those with the same name) has reduced 54,000 people labels to approximately 32,000.
- The result is a dataset of c. 160,000 letters. 32,000 people, 2,000 places
- The dataset has been annotated with additional information (Linked Open Data using WikiData/Wikipedia, group membership, e.g Royal Society, gender.
- Also included are abstracts - descriptions of the letters, from the Calendars



The Disambiguation Engine

Show whole list

39837 Samuel Hanson 39837 - -- fl. 1690. high sheriff of Carnarvonshire

39838 Samuel Harsnett, Archbishop of York 39838 link -- Samuel Harsnett (or Harsnet) (June 1561 – May 1631)

39839 Samuel Hartlib 39839; 39839; 39839; 39839; 39840; 39840; 39840; 39840; 39840; 0; 0; 0; 0 mapping: 39839; 3983 linked data: https://en.wik notes: --v1 Samuel H

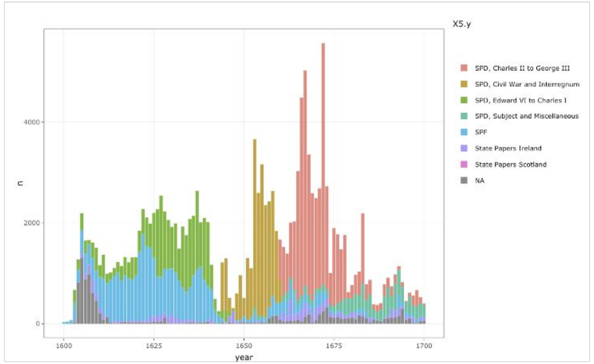
39840 Samuel Hartlinb 39840 - -v1 for Samuel Hartlib junior. evr this HAS to

1. SP 16/463 f.225
2. SP 16/463 f.227
3. SP 16/463 f.228
4. spo3 cspd c1 1640-1640 cspd c1 1640-1640-m1871-cm
5. SP 29/53 f.42
6. SP 29/164 f.99
7. SP 29/168 f.167
8. SP 29/173 f.34
9. SP 29/247 f.14
10. spo3 cspd com 1654-1654 cspd com 1654-1654-m1752x0073-cm
11. spo3 cspd com 1654-1654 cspd com 1654-1654-

Document ID: 4325301872
Image 1 | Image 2 | Image 3
Place: London.
Author: Samuel Hartlib
Recipient: Sir Thomas Roe
Document Ref.: SP 16/463 f.225
Title: Samuel Hartlib to Sir Thos. Roe.
Day: Aug. 10
Year: 1640

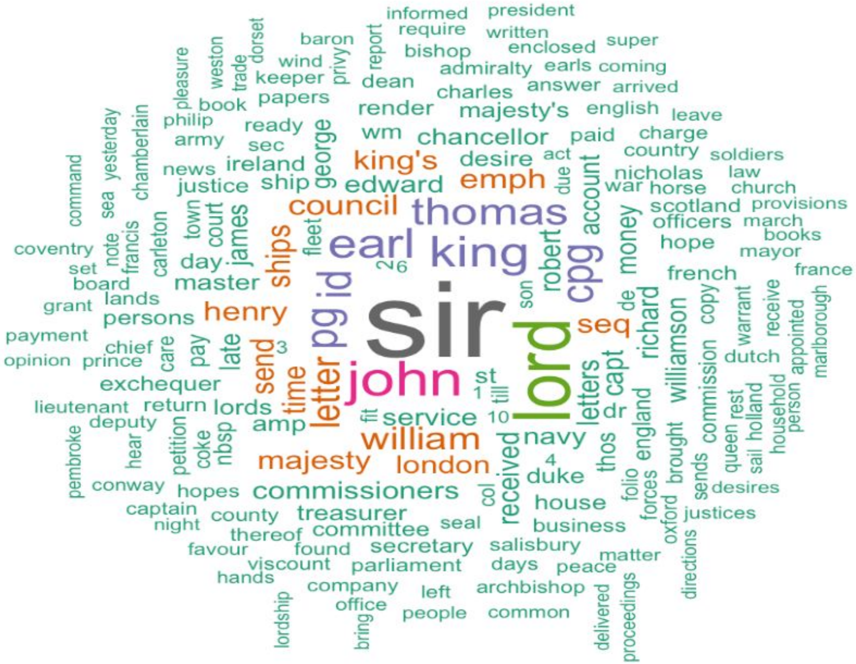
Samuel Hartlib to Sir Thos. Roe. These are to improve your leisure which perhaps you may enjoy in the retiredness from this place. The author of the Schedule of Divers New Inventions is the same Plattes who a year ago published two profitable treatises concerning husbandry and mines. He is

15 — Hartlib 0 spo3 cspd com 1652-1653 cspd com 1652-1653-m1131x103-cm
107 6 under clerks 0 spo3 cspd com 1658-1659 cspd com 1658-1659-m1426x0041-cm
364 Adam Saml. Hartman and Paulus Cyrillus 0 spo3 cspd com 1657-1658 cspd com 1657-1658-m1104x0046-cm SP 25/78 f.390
731 Alex. Hope 0 spo3 cspd com 1659-1660 cspd com 1659-1660-m1301x0375-cm
8182 Colonel Cressy Dimoke 8182 SP 29/295 f.67
10904 Crescy Dymocke 10904 SP 16/461 f.4 SP 16/461 f.8
12314 Dr. Benjamin Worsley 12314 SP 29/318 f.163 SP 29/337 f.239 SP 29/360 f.277
14569 Edward Savage 41318 SP 17/D f.17 SP 16/423 f.170
20065 Inhabitants of Dover 20065 SP 25/108 f.71
23382 John Holland 23382; 23382; 23382; 23382; 23382; 23382; 23382; 0; 0; 23382; 23382; 23382; 23382; 23382; 56410 (15 manuscripts)
24984 Joseph Avery 24984 SP 16/490 f.54 SP 63/301 f.172 SP 81/53/1 f.26 SP 88/9 f.76
39696 Sam. Hartlib 0 SP 25/71 f.126 spo3 cspd com 1654-1654 cspd com 1654-1654-m1752x0120-cm SP 25/105 f.84 SP 25/108 f.71



How can we ‘read’ the abstracts?

- A large dataset (174,264 documents, 24,331,502 tokens (words))
- Difficult to pinpoint interesting letters, besides the historically obvious ones
- Computer code can help to ‘read’ the abstracts in different ways, for example text mining, network analysis, or Named Entity Recognition



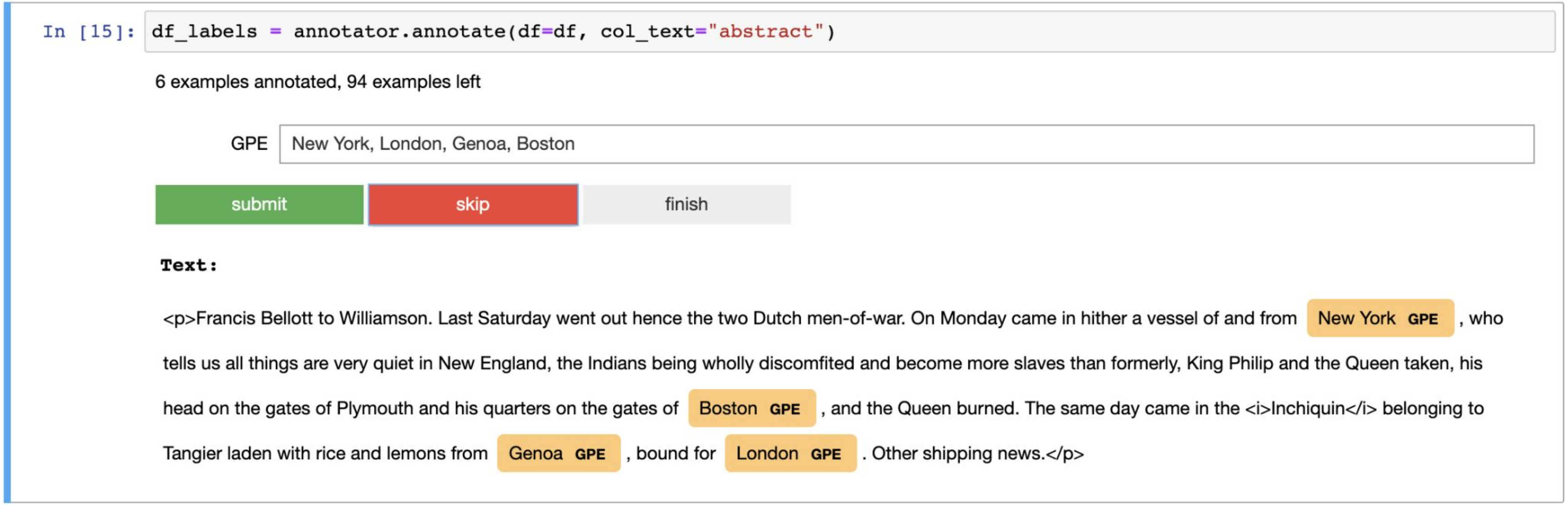
Extracting Geographic Information using Named Entity Recognition

- Using a statistical model, detect different types of ‘entities’ found in a piece of text.
- For example people, places, organisations..

Silas Taylor to Williamson. The generals are sailing in the Royal James, with the fleet, from the Weelings to the Texel; the Royal Charles is sent back to be repaired. Sir Win. Penn is sailing in the Henrietta towards the buoy of the Nore. Suggests the removal of Capt. Dorrell's company of the Admiral's regiment out of town, or into tents; Harwich is only part of a parish, with a chapel-of-ease, and has already quartered 200 shipwrights, besides calkers, labourers, &c., and many sick and wounded men are come or coming ashore.

Steps

- We used a Python library called SpaCy, but updated it with our own ‘training data’
- To do this we annotated 1,000 State Papers abstracts picked at random, marking the places found, then fed this back to the model.



```
df_labels = annotator.annotate(df=df, col_text="abstract")
```

6 examples annotated, 94 examples left

GPE: New York, London, Genoa, Boston

Text:

<p>Francis Bellott to Williamson. Last Saturday went out hence the two Dutch men-of-war. On Monday came in hither a vessel of and from New York GPE , who tells us all things are very quiet in New England, the Indians being wholly discomfited and become more slaves than formerly, King Philip and the Queen taken, his head on the gates of Plymouth and his quarters on the gates of Boston GPE , and the Queen burned. The same day came in the <i>Inchiquin</i> belonging to Tangier laden with rice and lemons from Genoa GPE , bound for London GPE . Other shipping news.</p>

Evaluating the model

- This gave us significantly improved results as measured by the number of places found out of the total it should have (precision), the number of places that were correct out of the total found (recall), and the average between the two (f-score).

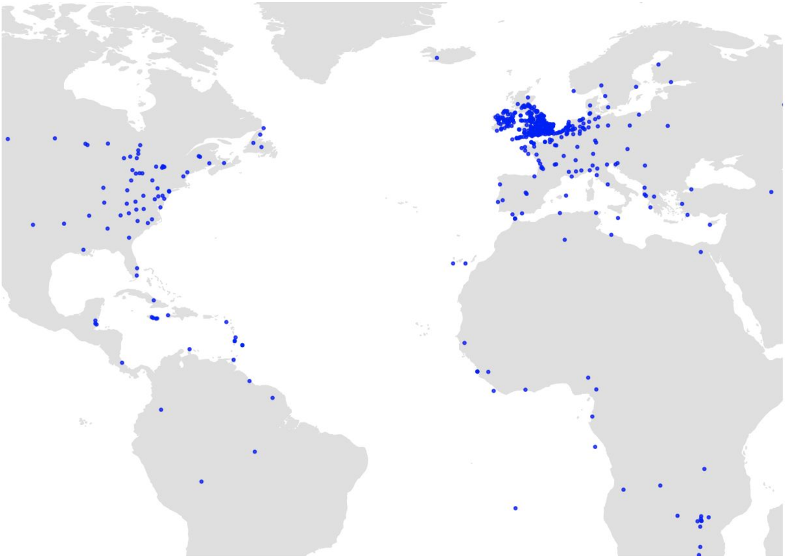
	Before Training	After training
Precision	81.797	86.002
Recall	43.794	94.146
F-Score	57.045	89.890

Geo-resolving

- SpaCy NER can detect geographic entities, but it doesn’t link them to coordinates on a map
- A second stage in the process was to build a custom geo-resolver.
- We made a gazetteer from three sources:
  - All the place names found in the State Papers Data (including variants and mistakes)
  - Place names found in the other dataset, EMLO
  - A database of world placenames, ‘Geonames’, which contains about 27,000,000 places.
- Then an algorithm to find the most likely match:
  - It checked each of the datasets in turn, prioritising them in order
  - If there were multiple matches, it took the match geographically closest

Results

- Haven’t run the model over the full dataset yet
- 1660–1680 (inclusive) detects about 195,000 entities (10,000 unique ones).
- It sometimes gets thrown by geographic titles e.g Duke of **York**
- The georesolver prioritises places local to the letter, which is not always the best guess, e.g **Boston** might be more likely to be in the U.S.
- Because the geonames dataset is so large, NER mistakes tend to get resolved to *somewhere,* often incorrectly.

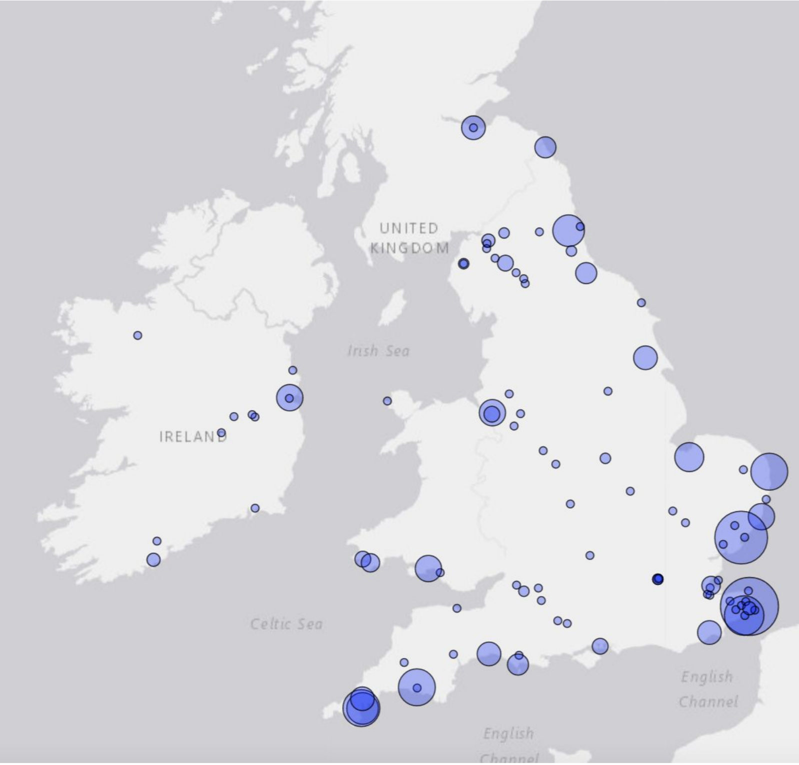


What to do with the data?

- Can be used as a dataset to analyse in its own right:
- Looking for ‘bursts’ of activity mentioning a certain place or geographic area
- Analyse the change in geographic focus of intelligencing over time
- Another way to navigate the archive:
- We built an application which allows the abstracts to be visualised and browsed based on places mentioned.

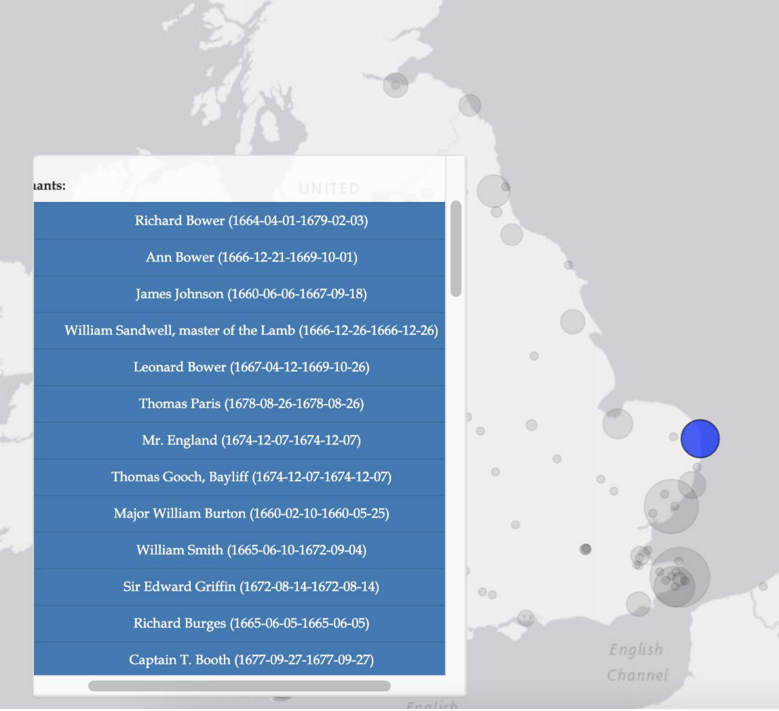
Navigating the archive with an interactive map

Default view is a map of all letters by place of sending



Navigating the archive with an interactive map

Clicking on a point shows all the senders

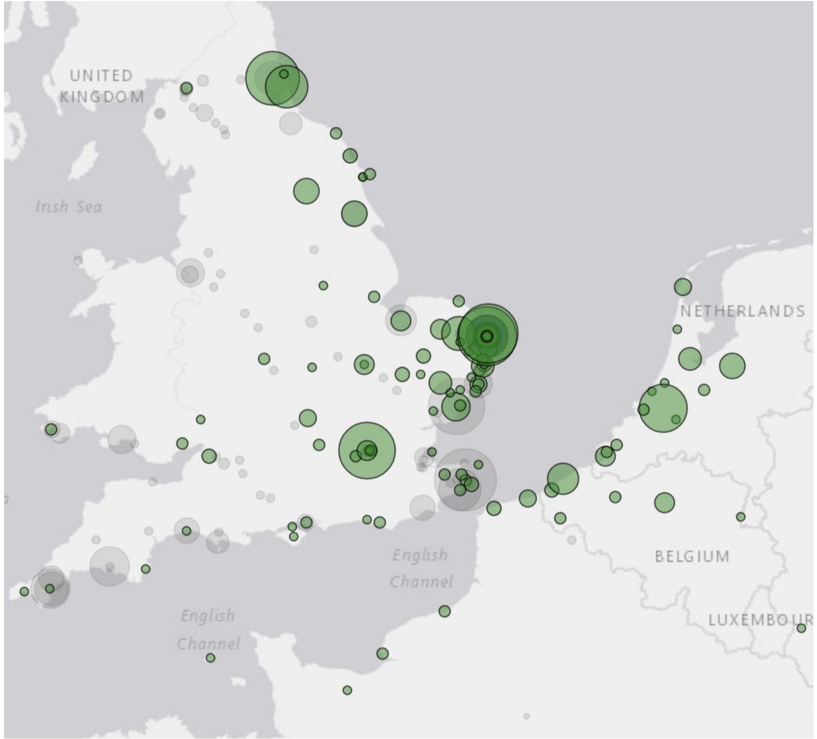


ıants:

- Richard Bower (1664-04-01-1679-02-03)
- Ann Bower (1666-12-21-1669-10-01)
- James Johnson (1660-06-06-1667-09-18)
- William Sandwell, master of the Lamb (1666-12-26-1666-12-26)
- Leonard Bower (1667-04-12-1669-10-26)
- Thomas Paris (1678-08-26-1678-08-26)
- Mr. England (1674-12-07-1674-12-07)
- Thomas Gooch, Bayliff (1674-12-07-1674-12-07)
- Major William Burton (1660-02-10-1660-05-25)
- William Smith (1665-06-10-1672-09-04)
- Sir Edward Griffin (1672-08-14-1672-08-14)
- Richard Burges (1665-06-05-1665-06-05)
- Captain T. Booth (1677-09-27-1677-09-27)

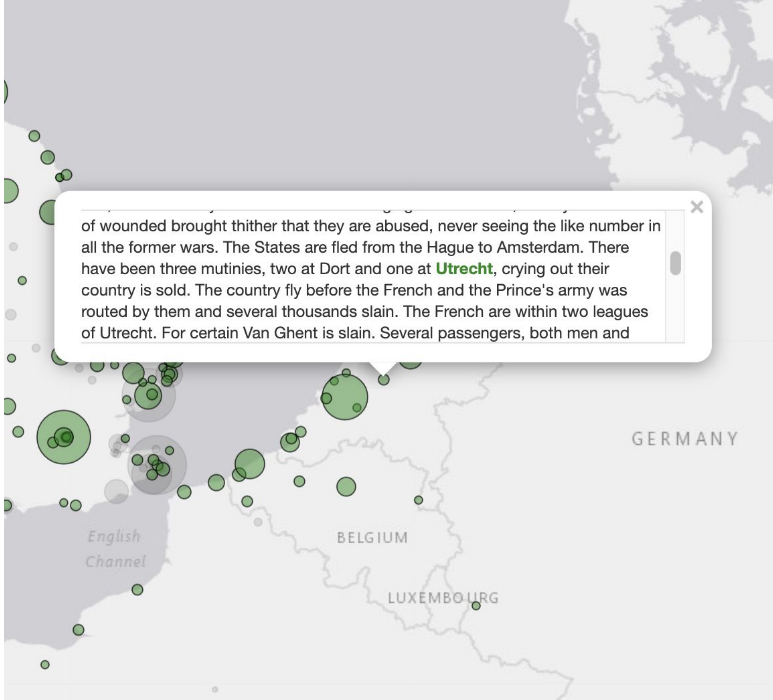
Navigating the archive with an interactive map

Clicking on a sender shows all the places mentioned



Navigating the archive with an interactive map

Clicking on one of these displays all the abstracts where that place is mentioned.



of wounded brought thither that they are abused, never seeing the like number in all the former wars. The States are fled from the Hague to Amsterdam. There have been three mutinies, two at Dort and one at **Utrecht**, crying out their country is sold. The country fly before the French and the Prince's army was routed by them and several thousands slain. The French are within two leagues of Utrecht. For certain Van Ghent is slain. Several passengers, both men and

Brief case study: News in the State papers

Letters of news and ‘intelligence’ are one of the most common types of exchange in the SP.

Some are notable but also tens of thousands of quotidian, repeating reports.

NER can help us to also ‘read’ these.

Ri. Watts to Williamson. The Earl of Sandwich has gone for Holland; five of the King's ships from Portsmouth and 17 merchant ships have arrived.

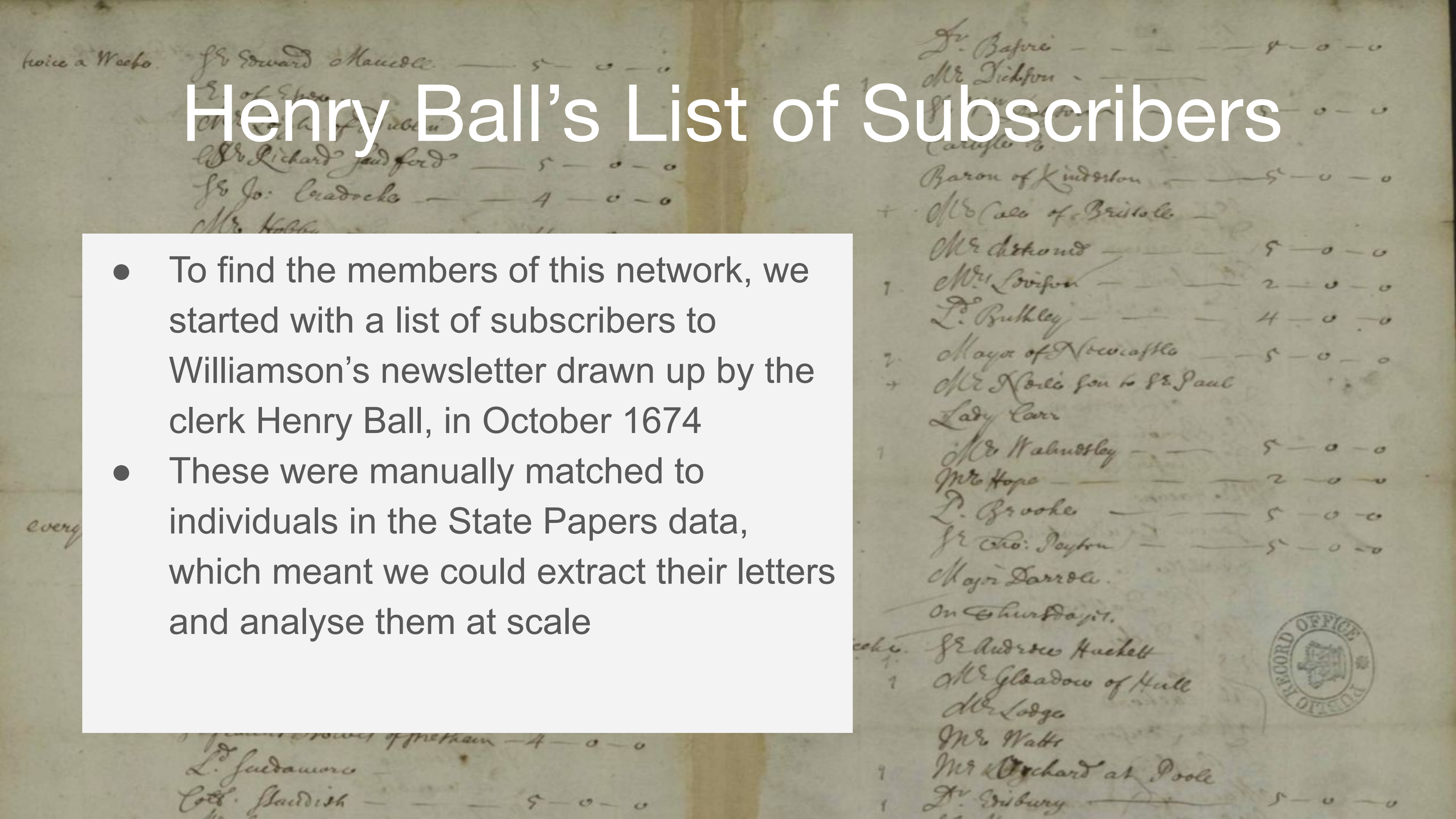
Ri. Watts to [Williamson]. Merchantmen, supposed to be Dutch, are sailing at the back of the Goodwin. The distemper decreases in Deal. Three quarters of those who stayed in the town are dead; it encreases in Sandwich. The Isle of Thanet and city of Canterbury are clear.

Hugh Salesbury to Williamson. The Constant Warwick has sailed for Ireland. All haste is made with the Revenge, Sir. Edw. Spragg being to command her. Sir Philip Honywood has got a surfeit by eating fresh salmon

Joseph Williamson (1633–1701)

- Sent or received almost 20,000 letters, from 2,000 correspondents. Largest number of letters and connections in SPO.
- Under-Secretary of State in 1660s and 70s, and superintendent of the London Gazette
- Williamson was ‘de facto head of a government intelligence system’
- Together with James Hickes he ran a newsletter service/news gathering operation

Henry Ball’s List of Subscribers



- To find the members of this network, we started with a list of subscribers to Williamson’s newsletter drawn up by the clerk Henry Ball, in October 1674
- These were manually matched to individuals in the State Papers data, which meant we could extract their letters and analyse them at scale

twice a Weeke. Sr Edward Mansell — 5 — 0 — 0
Earl of Essex
... of Dublin
Sr Richard Sandford — 5 — 0 — 0
Sr Jo: Cradocke — 4 — 0 — 0
Mr Hobbs

every

... Bowes of Stretham — 4 — 0 — 0
Ld Scudamore
Coll. Standish — 5 — 0 — 0

Dr Basire — 4 — 0 — 0
Mr Dickson
... — 0 — 0
Carlisle
Baron of Kinderton — 5 — 0 — 0
+ Mr Cale of Bristoll
Mr Ashond — 5 — 0 — 0
1 Mr Lowson — 2 — 0 — 0
Ld Bulkley — 4 — 0 — 0
1 Mayor of Newcastle — 5 — 0 — 0
+ Mr Nolle son to Sr Paul
Lady Carr
1 Mr Walmsley — 5 — 0 — 0
Mr Hope — 2 — 0 — 0
Ld Brooke — 5 — 0 — 0
Sr Tho: Peyton — 5 — 0 — 0
Major Darrell.
On Thursdays.
...ke. Sr Andrew Hackett
1 Mr Gleadow of Hull
Mr Lodge
Mr Watts
1 Mr Orchard at Poole
1 Dr Ewsbury — 5 — 0 — 0

PUBLIC RECORD OFFICE

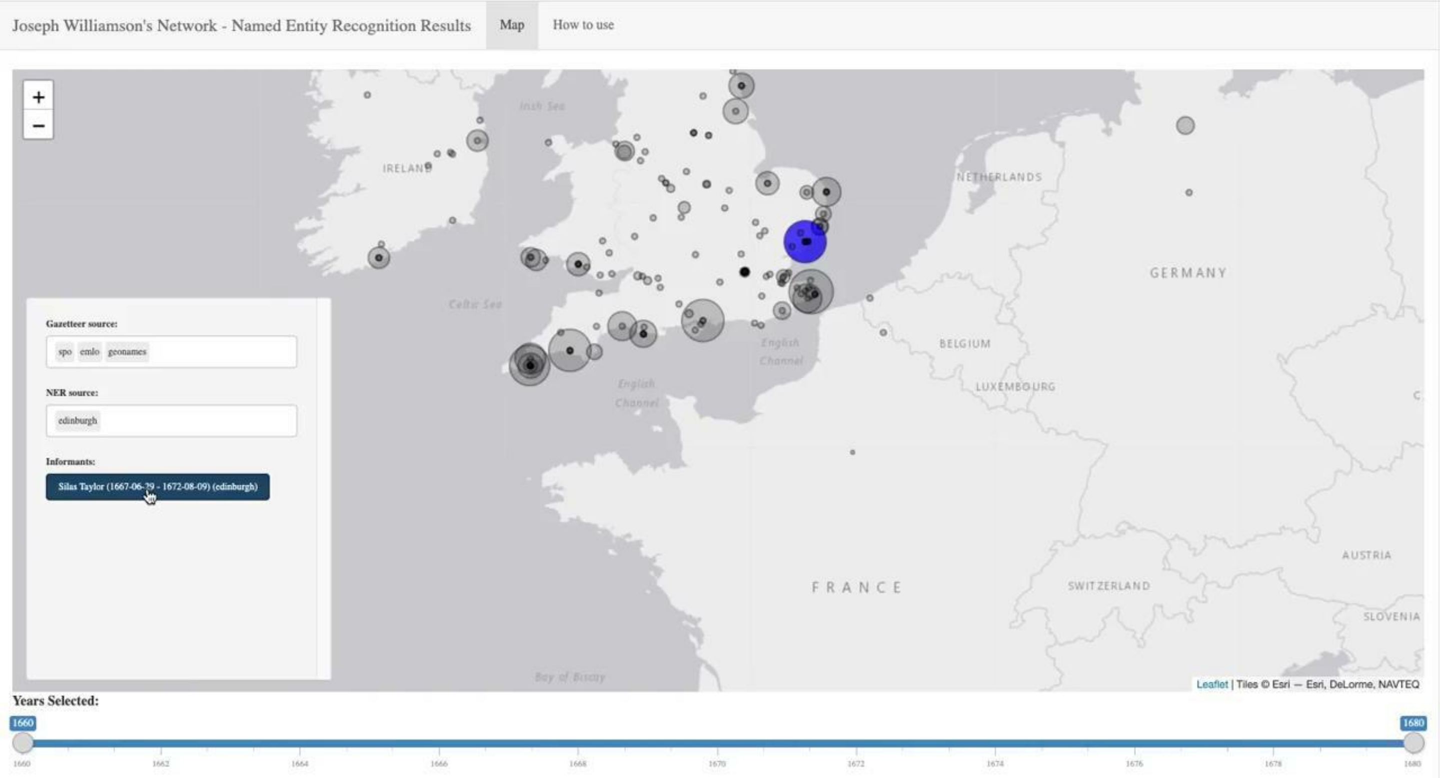
Mapping the Newsletter informants using NER

- We set out to understand more about the shape and extent of Williamson’s news network using DH techniques:
  - Identifying letters of news
  - ‘Reading’ the letters using Named Entity Recognition
  - Analysing and mapping the results

- We found many informants had a large, recognizable area of ‘specialism’, and for the coastal informants who make up the bulk of the letters, this was a wide area determined by the ports from where ships were arriving to that town.

- The tool can help identify individual letters of interest and trace the movement of news

The Tool in Action..



Joseph Williamson's Network - Named Entity Recognition Results

Gazetteer source: spo emlo geonames

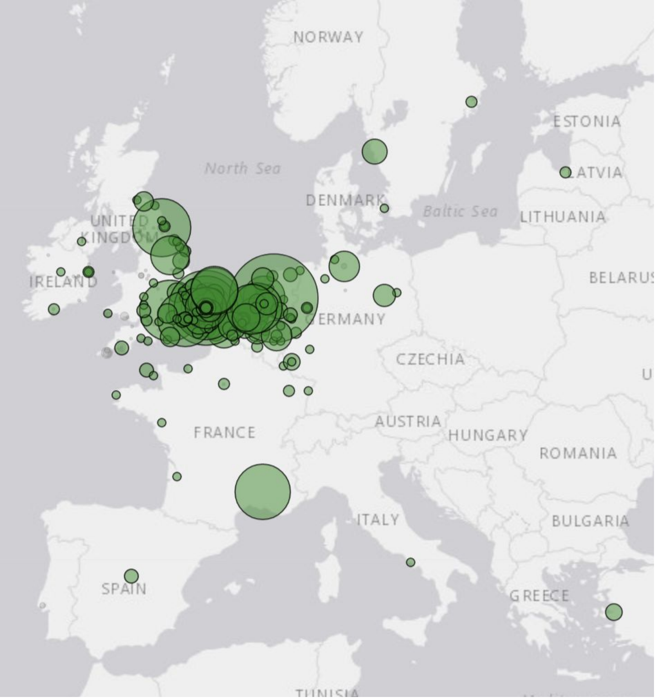
NER source: edinburgh

Informants: Silas Taylor (1667-06-29 - 1672-08-09) (edinburgh)

Years Selected: 1660 – 1680

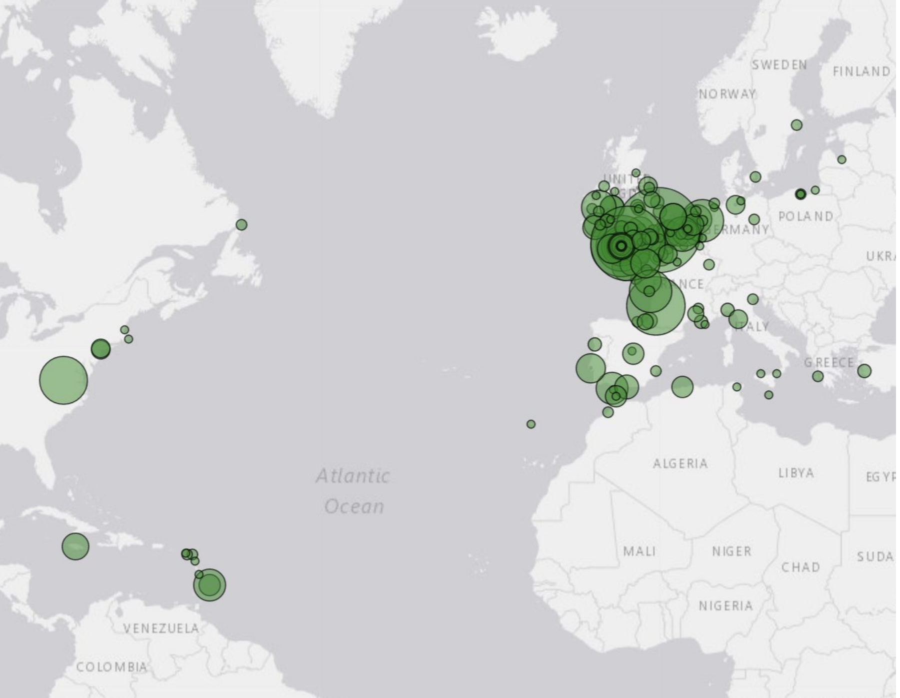
Silas Taylor, Harwich

- One of the largest subsets of letters in the State Papers
- Mostly domestic, south-east coast, plus Low Countries
- Received news from an international network:
    - Silas Taylor to [Williamson], Yesterday I discoursed with a very understanding person from **Hamburg**, who within this fortnight received a letter from a Scotch colonel at **Berlin**, wherein he, taking notice of the reports of succours from the Dukes of Brandenburg and Lünenburg, avers that there is not the least motion of a march towards Holland, nor any forces there beyond what are usually kept up.
- (We can also see errors, such as where the ‘Orange’ in ‘Prince of Orange’ has been mapped to the location in the South of France).



Thomas Holden, Falmouth

- Wider area of news, due to shipping connections to the Atlantic and Mediterranean
- Sends Williamson news from America and the Carribbean:
    - Thomas Holden to Williamson. The 19th came in the William and Mary from **Barbados**, after a passage of eleven weeks, from contrary winds and foul weather. About twenty more sail were almost ready when she came away. The 20th came in the Providence and the 21st the Elias and the Rebecca, all from **Virginia**, laden with tobacco. They all speak of the thriving condition of that place and of the goodness of the year. Wind now S.W.



What does it tell us about Williamson’s network?

- This shows the importance of these coastal informants: much news from Europe was coming from coastal towns and being sent to London - challenges the idea of ‘centre’ and ‘periphery’
- Not just reliant on typical historical sources (post office records, accounts and so forth) to understand the geography of the news network but can use the letters of news themselves as a source
- Williamson’s foreign intelligence was at least in part supplied by these ordinary, part-time news writers.

What does the NER help us to do?

- Make sense of a group of letters which are often overlooked, ie. the everyday letters of news which make up a large part of the SP archive
- A new way to navigate the State Papers and highlight items of interest - not bound to browsing by date or series numbers, or by specific keyword searches
- ‘Reshuffule’ the organisational structure of the archive as it stands
- Understand more about the geography of the early modern state and in particular the geographic patterns of intelligencers.

Next steps:

- Improve the NER model with annotations taken from a wider sample of the State Papers, evaluate its use on other letter datasets
- Use more contextual information to geo-resolve NER results
- Carry out some of the analyses mentioned earlier (detecting events through ‘bursts’ of place mentions
- Eventually build a model for full-text transcriptions or outputs from Handwritten Text Recognition